VOL. 6 No. 1, pages 1-32 JANUARY 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

THE NEW BROADCASTING TRANSMITTERS IN THE NETHERLANDS

by H. B. R. BOOSMAN. 621.396.712

Two new broadcasting transmitters of 125 kW have recently been constructed to replace the Netherlands broadcasting transmitters in Hilversum and Jaarsveld. Both the new transmitters are housed in the same building. This article gives a general account of the installation.

At the beginning of this year the Netherlands broadcasting transmitters of Hilversum and Jaarsveld were replaced by a single central installation, which is so situated that it gives optimum reception throughout the whole of the Netherlands. This new installation, supplied by Philips, consists of two transmitters, each of 125 kW, which work on 336 m and 415 m, respectively. The wave length of 336 m was assigned to the Netherlands by the Montreux plan; at present this plan is not in operation, and therefore the transmitter in question is temporarily tuned to a wave length of 301.5 m.

In designing the new transmitters special attention was paid to the efficiency, in connection with the high power to be radiated. The necessity of achieving high efficiency led to connections with which the transmission energy is supplied to the aerial through four almost identical channels, each consisting of three successive stages. Compared with normal connections, in which the modulated high-frequency energy is amplified in a series of successive stages to the desired level, an economy of more than 250 kW has been obtained in this way.

It is not the intention of this article to describe the action of the connections with four channels, or to go into the way in which the improvement in the efficiency is achieved. This will be dealt with in a future article in this periodical. In this article, however, we shall discuss the influence which the splitting of the transmitter connections into four channels has had on the design of the whole installation, after which several particulars of the construction will be considered.

The general plan

The most obvious plan for an installation on the principle mentioned is represented in *fig. 1*, where the fanshaped design meets the necessity of larger dimensions for each successive stage. As may be seen the splitting of the system into four channels does not take place immediately, but in two steps, upon transition from the second to the third stage and from the third to the fourth stage. The fourth stage is controlled by the modulator, and each of the channels is affected in a different way. The resulting unequal division of the transmission energy

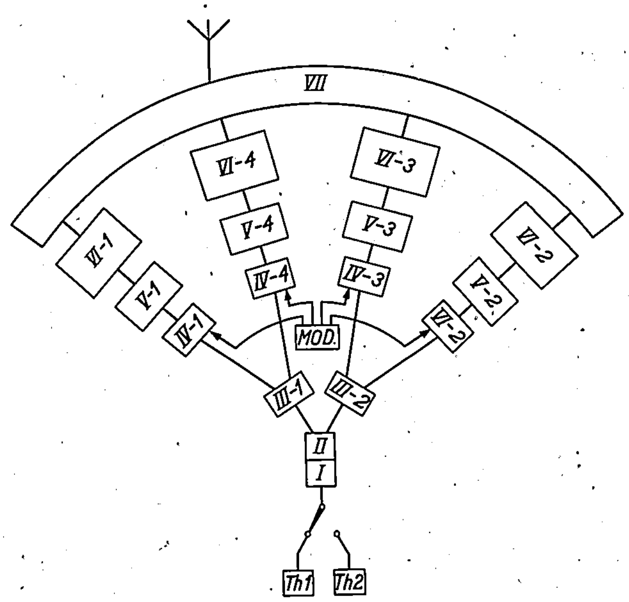


Fig. 1. Scheme of the new Netherlands broadcasting transmitters, represented in a form which could be considered as a ground plan. The high-frequency energy is amplified in six stages (*I-IV*), of which stage *III* is split into two channels and stages *IV*, *V* and *VI* into four channels. *VII* is a network which couples the four output stages with each other in a certain way, and transmits the energy provided to the aerial. *MOD* is a modulator acting on stage *IV* which regulates the amplitude of the high-frequency signal in the four channels, in a special way for each channel. *Th 1* and *Th 2* thermostats with quartz crystals for keeping the frequency constant.

is the means by which the high efficiency is attained.

The fan-shaped design has, in principle, the advantage that the four channels could with little difficulty be built so that they may be considered mutually identical, which is desirable in order to prevent relative phase shifts between the high-frequency signals of the four channels. Such phase shifts are not permissible if the satisfactory functioning of the whole unit is at stake. Practically, however, this design is less satisfactory and was finally abandoned since the operation as well as the control become fairly difficult.

Ease in operation and control requires as compact as possible an arrangement, which can be realized for the first four stages, together with the modulator, in the manner shown in *fig. 2a*. The sixth stage, which, like the fourth and fifth, consists of four channels, requires more space, and it was therefore set up separately (see fig. 2*b*). The coupling of stage *IV* with stage *V* and of stage *V* with stage *VI* is by means of connection lines which are of exactly the same length for the different channels. The identity of the four stages as to the phase of the high-frequency signal is sufficiently well guaranteed by this means.

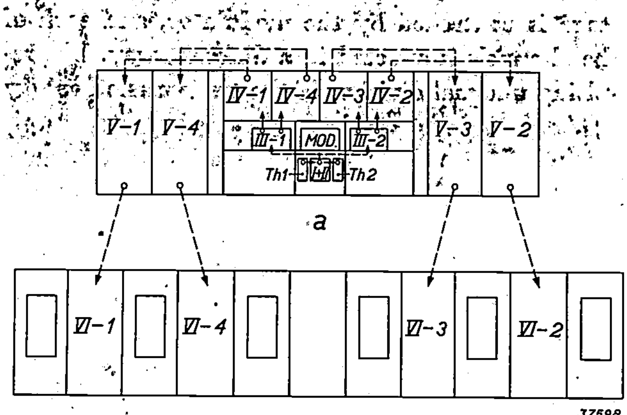


Fig. 2. Diagrammatic front view of the transmitter as installed.

The relative arrangement of the parts of the transmitter *a* and *b* corresponds exactly to the sketch just discussed. This is not to be considered as a ground plane, but as a vertical cross section, *i.e.* the preliminary stages *a* are situated on a gallery above the final stage *b*. The whole transmitter is set up against one wall of a large hall; along the opposite wall is the second transmitter, which corresponds exactly to the first except for its wave length.

Fig. 3 represents a transverse cross section through the middle of the transmitter building, which consists in the main of two storeys, on the ground floor the machine hall (*I*) and above it the transmitter hall (*II*). In the machine hall are the different feed sources of the installation: for low voltages and heavy currents converters are installed, while the high voltage of the final stage is obtained with the help of a transformer and rectifier. On the galleries of the machine hall stand the insulation spirals which provide the necessary length of path for the cooling water to prevent the high potential of the anodes of the output valves being conducted away to earth by way of the cooling water.

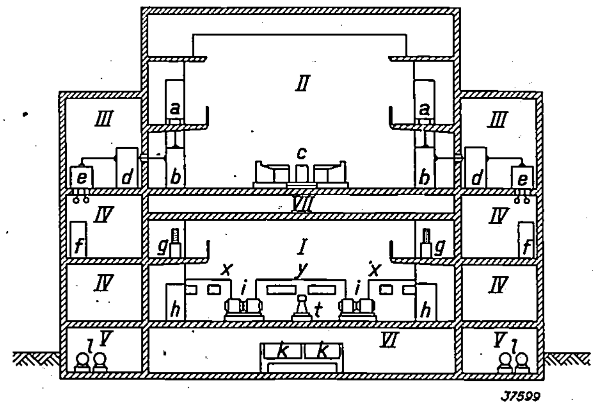


Fig. 3. Vertical cross section through the transmitter building.

I	Machine hall	*d*	anode circuits of output stage
II and *III*	Transmitter hall	*e*	aerial coupling circuit and filters
IV	Workshop, etc.	*f*	artificial aerials
V	Pump cellar	*g*	cooling-water spirals
VI	Cable cellar	*h*	switch board machine room
VII	Cable space	*i*	converters
a	preliminary stages	*x*, *y*	20 kV rectifiers
b	output stage		
c	operation desk		

Fig. 4 is a ground plan of the machine hall. *x* and *y* are rectifiers for the high voltage of the final stages; the transformers *v* and *w*, respectively, belong to these rectifiers. Each of the two rectifiers *x* serves for the supply of one transmitter, while *y* forms the common reserve. The converters for the low voltages are indicated by *i*. Three heating-current dynamos are used, namely, of 15, 25 and 35 volts, respectively, for 50, 320 and 830 amp., in order not to be compelled to dissipate too much energy in series resistances for the filaments of lower voltage by the use of one dynamo of 35 volts. Furthermore a double dynamo for 2 000 volts, 2 amp. and 450 volts, 1 amp. furnishes the anode voltages and screen-grid voltages of various preliminary stages, while two double dynamos, each for 400 volts, 2.5 amp. give the grid bias voltages for the four channels of the final stage. Finally there is a dynamo for 110 volts, 18 amp. for the excitation of the other machines. The motive energy for the dynamos is taken from the local high-voltage mains with the help of the transformers *r*, which are built into

separate cells at the end of the machine hall. Here also one transformer serves each transmitter normally, while a third transformer is in reserve.

For the further arrangement of the machine hall we refer to the text under fig. 4. Finally in *fig. 5* two photographs of the machine hall are reproduced: in one photograph may be seen the middle rectifier (y in fig. 4) behind a set of converters, while the other shows the switch board for the feed voltages (h in fig. 4).

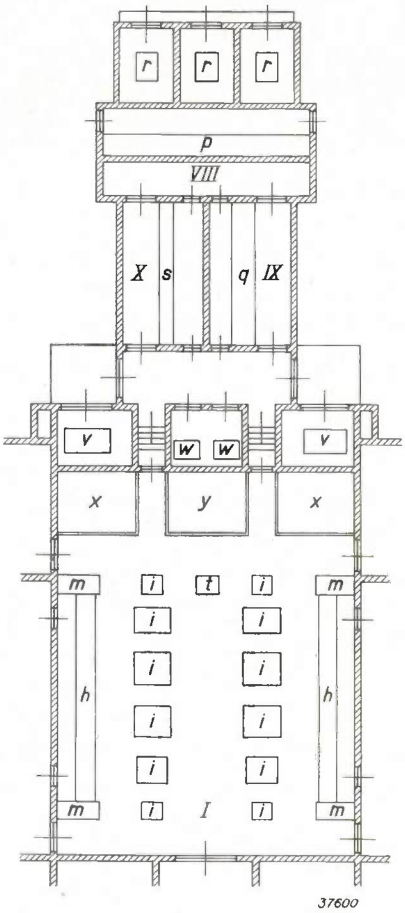


Fig. 4. Ground plan of the machine hall

I Machine hall — *m* cable channels
VIII Cable switch room — *r* mains transformers
IX High-voltage switchroom — *w* transformers for 20 kV rectifiers
X Low-voltage switch room

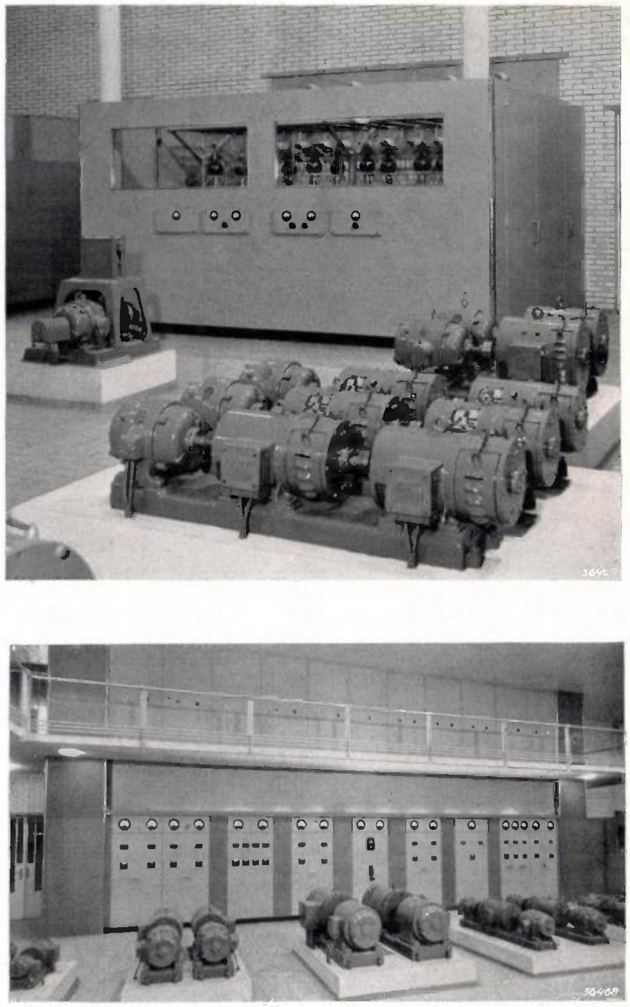


Fig. 5. Machine hall. Upper photograph, the 20 kV rectifier behind a set of converters; lower photograph, the switch board belonging to the converters.

As already mentioned, the transmitters proper are in the hall above the machine hall. The plan is given in *fig. 6*. The different components of the transmitter are indicated in the same way as in fig. 3, so that the sketch is clear without much explanation. At the front are two rooms, one for the head of works (*XI*) and one for an amplifier room (*XII*), with a view through broad windows into the transmitter hall. In the amplifier room the microphone amplifiers are situated on a rack (*O*). Along the side walls of the hall are the final stages (*b*) of the four channels; above on the galleries, not

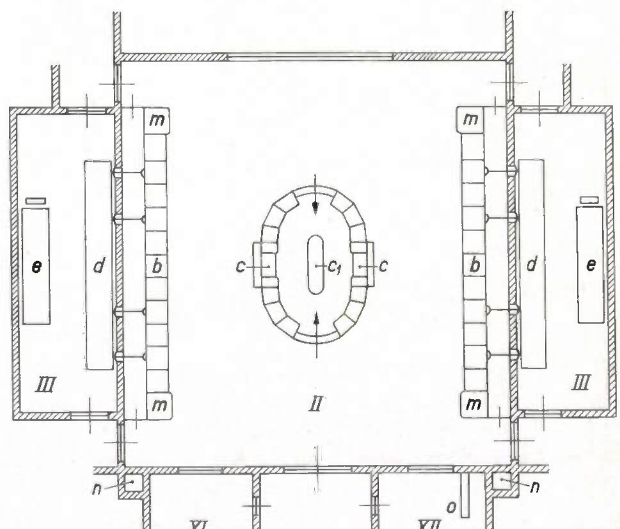


Fig. 6. Ground plan of the transmitter hall and the adjacent rooms. For the meaning of the letters and numbers see fig. 3 *XI* room for the works manager, *XII* amplifier room with amplifier rack *O*.

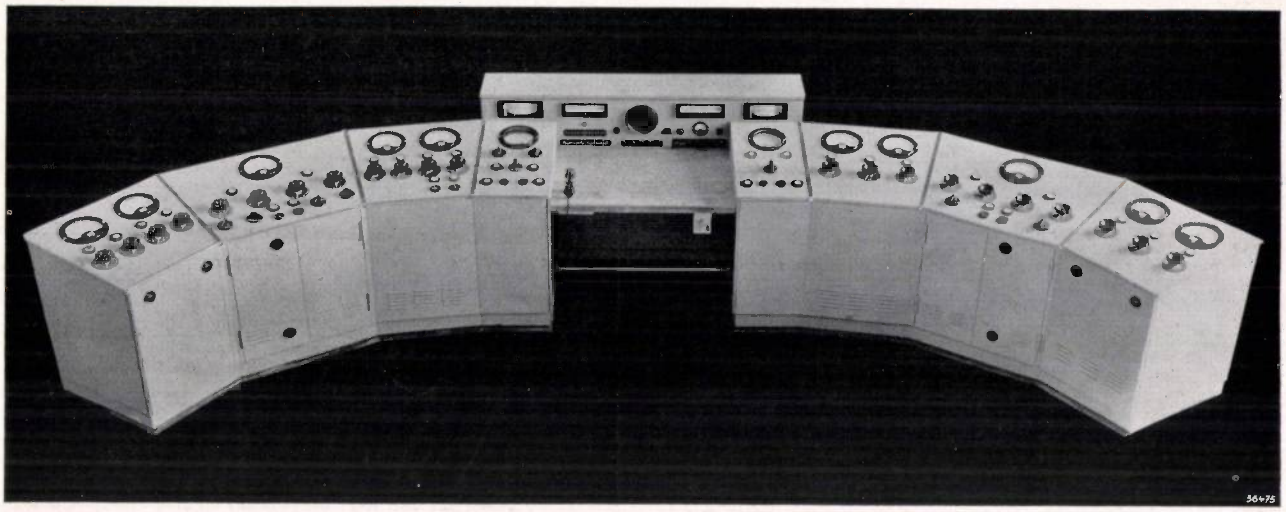


Fig. 7. Supervising and operating desk.

given in the floor plan, are the preliminary stages. The cables and cooling-water lines to these perliminary stages run through the cable channels (m) which may be seen on either side of the final stages. In the side rooms *III* of the transmitter hall is the anode network d of the final stage, which couples the four channels with each other, and also a cabinet e with the coupling circuit which transmits the output energy of the transmitter, after the higher harmonics have been filtered out, to the aerial. In

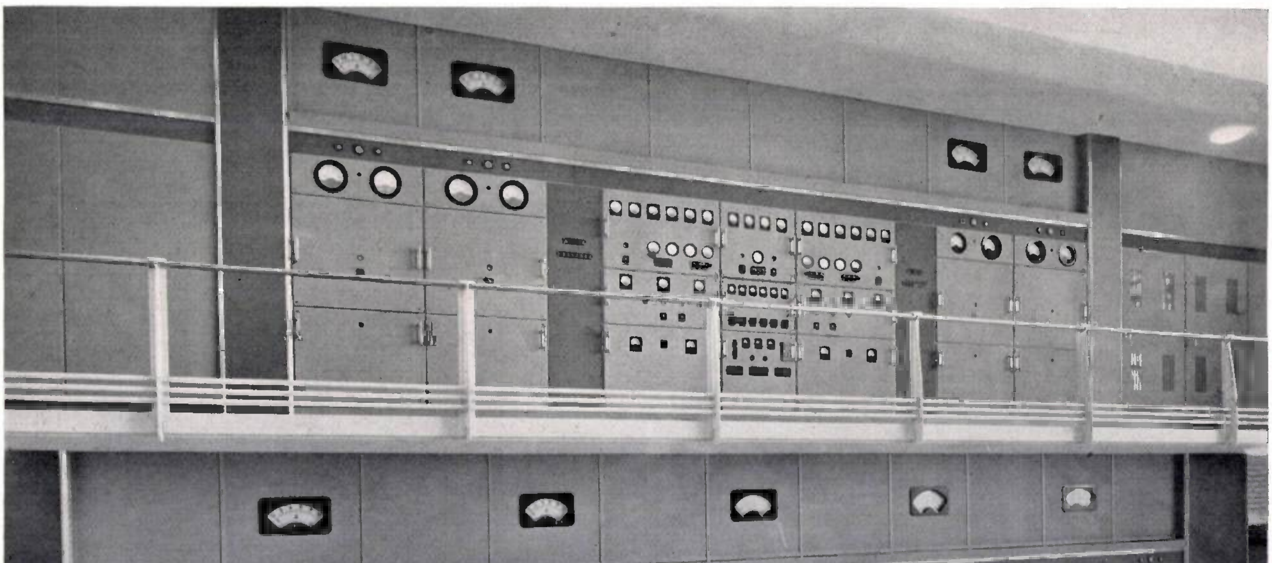


Fig. 8. Front of the transmitter seen from the operating desk. On the gallery the preliminary stages *I-V*: below the output stage *VI*. The sub-division of the front panel follows from the diagram of fig. 2.

the middle of the hall is the operating desk, half of which (for one transmitter) is shown in *fig. 7*. All the manipulations necessary for operating both transmitters can be carried out from this desk, with the exception of the switching in and over of the machines, which takes place in the machine hall itself, with simultaneous control by means of the measuring instruments present there. In *fig. 8* the panel of the whole transmitter has been photographed from the desk to show how, thanks to the arrangement in two storeys and the generous dimensions of the instruments, a good view of the whole installation is obtained. It is also striking that no operating knobs occur anywhere on the transmitter panel. This has been expressly avoided in order to prevent disadjustment of the transmitter, for instance by careless visitors. All the operating shafts can be moved by the personnel with the help of a removable handle.

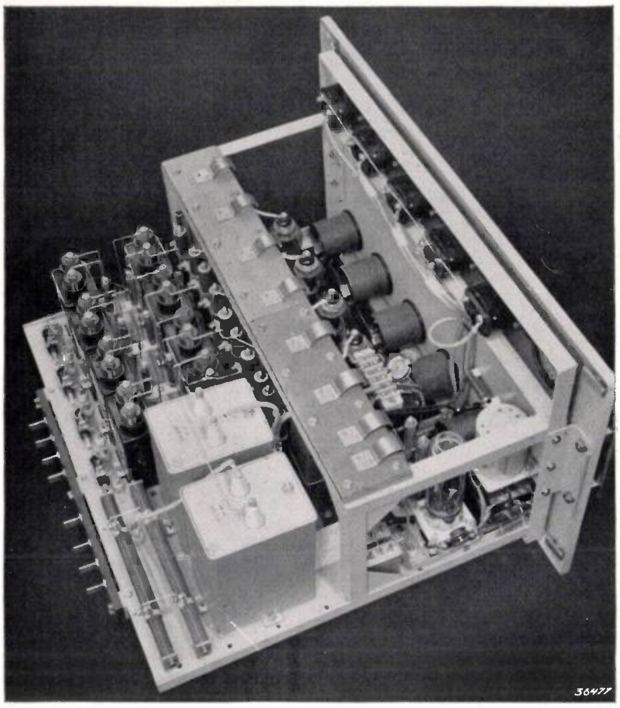


Fig. 9. Sliding chassis containing the modulator stage. To the left may be seen the knife contacts which serve for the electrical connections when the chassis is in position.

Several structural details of the transmitter

In the construction of the transmitter, spaciousness has been aimed at, in order that all components should be easily accessible. This is particularly important for those components which must be able to be exchanged, such as valves and fuses.

In order to facilitate access to the components the front panels of the transmitter are fastened with toggles and provided with handles, and all the valves are set up behind doors. Smaller components are mounted in sliding chassis (see *fig. 9*) with the employment of knife contacts for the passage of the current.

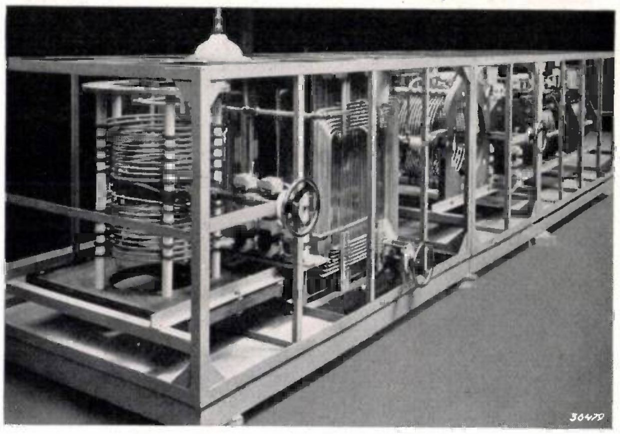


Fig. 10. Coupling circuit which transmits the energy from the output stage to the aerial. Also filters for eliminating the higher harmonics from the aerial signal.

A further discussion of the electrical construction of the transmitter lies outside the scope of this article. In order, however, to give some idea of the mechanical construction of the electrical components, three photographs of characteristic compo-

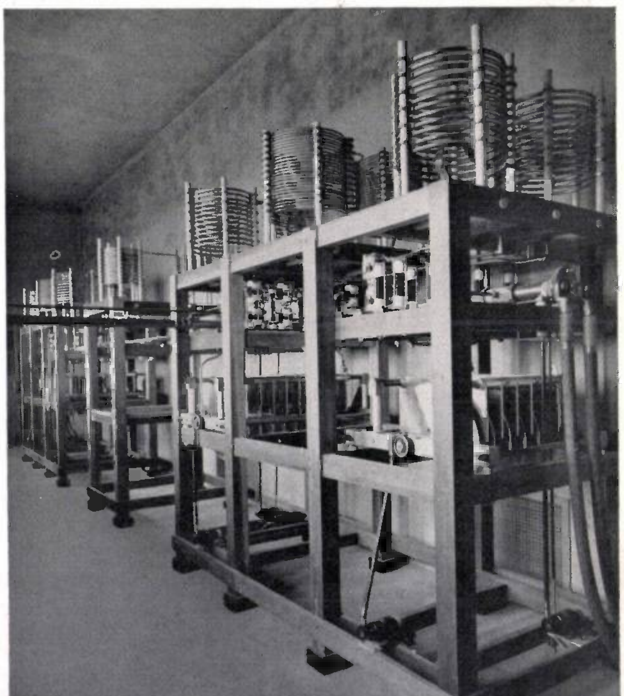


Fig. 11. Tuning circuits for the output stage of the transmitter. Below the astatically arranged coils may be seen the tuning condensers which are operated by motors.

nents are given in *figs. 10, 11* and *12*. We wish to call special attention to the construction of the coils of the high-frequency output stages in fig. 11. These coils are subdivided into two astatic halves, so that the mutual coupling of the coils is kept extremely small without external shielding.

Two self-radiating masts of 192 m and 165 m, respectively serve as aerials for the transmission waves of 415 m and 356 m. The earth connections network consists of 108 copper wires of 200 m and 180 m length, respectively, buried in the ground in ray arrangement between an inner and an outer ring.

Tuning of the transmitter

Except in the final stage, the circuits of the transmitter can be tuned in quite the ordinary way, since the preliminary stages of the different channels are independent of each other. Since, however, the channels of the final stage act on a common anode circuit, special measures had to be taken for the tuning of the final stage. The network which couples

Fig. 13. Measuring wagon with cathode ray oscillograph for supervising the relative phases of the voltages in the four channels.

the anode circuits of the output valves with each other is such that the anode circuit of one channel can most easily be tuned when the anode circuits of the adjacent channels are short circuited. The short circuiting is accomplished with the help of mechanical switches. These switches are operated with servo motors from the front of the transmitter, like the tuning condensers of the different channels. The anode voltage of the valves with short-circuited anode circuit can be disconnected by switches. These switches are electrically locked to the corresponding short-circuit switches, so that it is impossible to switch on the anode voltage of valves with a short-circuited anode circuit.

Fig. 12. Insulation spirals through which flows the cooling water for the anodes of the output valves. The water from six valves (four channels and two reserve valves) must pass through an insulation spiral both coming and going, so that 12 such spirals are needed. They are made entirely of porcelain and have a capacity of 120 litres of cooling-water per minute.

When in this way one channel is separated from the rest the oscillation circuits can be tuned, and by the adjustment of certain bias voltages in the modulation stage the carrier wave can be influenced in the desired way by the modulation voltage. As already stated, the nature of this influence is different for each channel; it is best to judge the relation between the modulation voltage and the amplitude of the output signal experimentally with the help of four cathode ray tubes which are built into the panel of the transmitter.

For satisfactory performance of the transmitter, not only are the amplitudes of the signals in the four channels important, but also the relative phases. In order to avoid phase shifts it is necessary to compensate exactly the grid-anode capacities of the triodes TA 18/100 used in the output stage, which vary slightly for different valves of that type. The neutrodyne condensers used for that purpose are constructed in two parts, the main part of which is adjusted once, while the correction part can be regulated from the front panel of the transmitter. The regulation can best be accomplished by checking the correctness of the relative phases of the four channels by means of a cathode ray oscillograph. For this purpose a transportable apparatus in the form of a measuring wagon is used. This is shown in *fig. 13*.

In *fig. 14* may be seen the rack and pinion drive by which the correction part of the neutrodyne capacity is adjusted; this picture is also characteristic of the way in which the tunable components housed in closed compartments are operated by means of sliding switches. The neutrodyning must be repeated every time a transmitter valve is replaced by a new one. In order to facilitate this a spare valve is built in for each two channels, which is previously neutrodyned with the help of

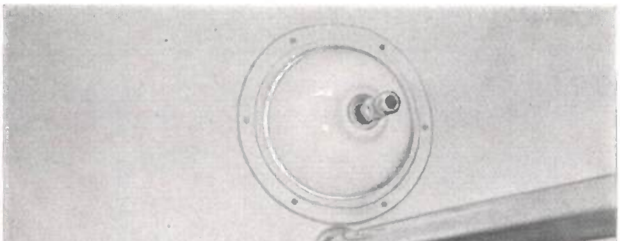


Fig. 14. Neutrodyne condenser which is set by means of a rack and pinion.

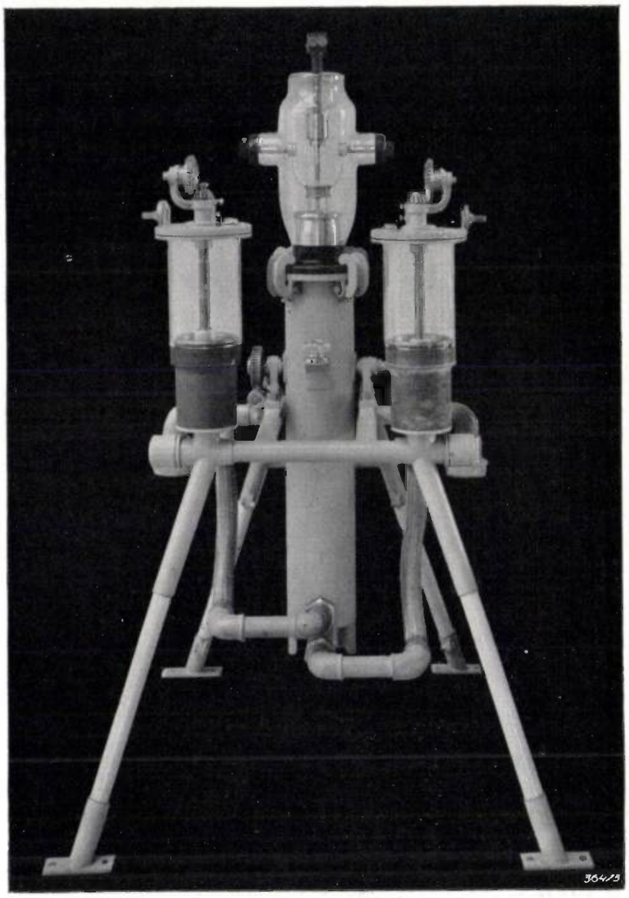


Fig. 15. Spare valve with two neutrodyne condensers, which are adjusted for the two channels, respectively, in which the valve there present might have to be replaced.

two neutrodyne condensers for each of the two channels (see *fig. 15*). Switching over to a spare valve therefore only requires a minimum time (about 15 sec) and can be done by unskilled personnel.

Characteristics of the transmitter

Compared with the ordinary transmitter connections, the installation described, in which the

Distortion of low-frequency signals at a modulation depth of 90%		Efficiency at different depths of modulation		
frequency c/s	distortion		efficiency output stage	efficiency of whole transmitter
300	3.2%	carrier-wave without mod.	58.9%	36%
400	3.0%			
2 000	3.9%	30% mod.	56%	35%
6 000	5.5%			
9 000	6.3%	100% mod.	55%	38%

modulation is distributed over four channels, offers the advantage of a considerably higher efficiency. The degree of efficiency depends upon the way in which this division is carried out. The higher the desired efficiency, the greater the distortions which must in general be tolerated; the final result may thus be considered as a compromise between the efficiency of the transmitter and the fidelity of the reproduction. In the case of the transmitter as adjusted for the acceptance examination the values for efficiency and distortion given in the above table were measured.

If, in agreement with practical experience, it is assumed that with the ordinary construction of the transmitter an efficiency of only about 20 per cent would have been attained, the present construction represents an economy of more than 250 kW. This economy is particularly important because it not only means an economy in the current consumption, but at the same time it has a favourable effect on the dimensions of the water-cooling system and other components which must dissipate the heat developed without harmful rise in temperature.

NEW KINDS OF STEEL OF HIGH MAGNETIC POWER

by B. JONAS and H. J. MEERKAMP van EMBDEN. 669.15.018.58

By exposing steel for permanent magnets to a magnetic field during the heat treatment (magnetic hardening), considerable improvement in the magnetic qualities can be attained in certain cases. The maximum product of induction and internal field strength (magnetic power), which determines the quality of a magnet steel, could in this way be increased from 2.2×10^6 to 5.2×10^6 gauss-oersted in the case of the magnet steel alloy "Ticonal". The significance of this progress is briefly explained in this article.

Although certain general ideas already exist about the magnetic properties of alloys, in working out the most favourable composition of an alloy of which certain magnetic properties are required, experience is still almost the only guide. This is true also for the alloys for permanent magnets which have been developed in the Philips Laboratory chiefly for use in electrodynamic loud speakers.

Although the starting point of these investigations was the carbon-cobalt type of steel then used for that purpose, after the discovery of Mishima in 1932 that better and cheaper permanent magnets could be made from iron alloys with nickel and aluminium in certain proportions, magnet steels were developed which contained these latter elements as well as titanium and cobalt [1]).

In the research on these new kinds of steel, which were sold under the name "Ticonal", not only was the composition varied in many ways, but also the nature of the heat treatment. The results were judged in each case by recording the magnetization curve and especially the remanence B_r and the coercive force H_c (see *fig. 1*). It was found that coercive forces of more than 1 000 oersted could be obtained. The remanence is usually found to decrease with increasing coercive force, so that the product of coercive force and remanence, which to a certain point may be considered as a measure of the magnetic power of the steel, cannot be increased above a certain limit of about 6×10^6 gauss-oersted.

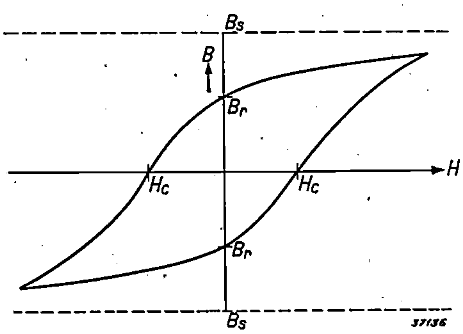


Fig. 1. Hysteresis loop of a magnet steel. B_r remanence, H_c coercive force.

Actually the quality of a magnet steel cannot be determined exactly by the product of remanence and coercive force, but by the maximum value $(BH)_{max}$ of the product of induction and internal field strength, which may occur when the material

[1]) On the fundamental difference between the magnetic hardening phenomenon in the case of these alloys and that of the earlier known kinds of magnet steel see J. L. Snoek, Philips techn. Rev. 2, 233, 1937.

modulation is distributed over four channels, offers the advantage of a considerably higher efficiency. The degree of efficiency depends upon the way in which this division is carried out. The higher the desired efficiency, the greater the distortions which must in general be tolerated; the final result may thus be considered as a compromise between the efficiency of the transmitter and the fidelity of the reproduction. In the case of the transmitter as adjusted for the acceptance examination the values for efficiency and distortion given in the above table were measured.

If, in agreement with practical experience, it is assumed that with the ordinary construction of the transmitter an efficiency of only about 20 per cent would have been attained, the present construction represents an economy of more than 250 kW. This economy is particularly important because it not only means an economy in the current consumption, but at the same time it has a favourable effect on the dimensions of the water-cooling system and other components which must dissipate the heat developed without harmful rise in temperature.

NEW KINDS OF STEEL OF HIGH MAGNETIC POWER

by B. JONAS and H. J. MEERKAMP van EMBDEN. 669.15.018.58

By exposing steel for permanent magnets to a magnetic field during the heat treatment (magnetic hardening), considerable improvement in the magnetic qualities can be attained in certain cases. The maximum product of induction and internal field strength (magnetic power), which determines the quality of a magnet steel, could in this way be increased from 2.2×10^6 to 5.2×10^6 gauss-oersted in the case of the magnet steel alloy "Ticonal". The significance of this progress is briefly explained in this article.

Although certain general ideas already exist about the magnetic properties of alloys, in working out the most favourable composition of an alloy of which certain magnetic properties are required, experience is still almost the only guide. This is true also for the alloys for permanent magnets which have been developed in the Philips Laboratory chiefly for use in electrodynamic loud speakers.

Although the starting point of these investigations was the carbon-cobalt type of steel then used for that purpose, after the discovery of Mishima in 1932 that better and cheaper permanent magnets could be made from iron alloys with nickel and aluminium in certain proportions, magnet steels were developed which contained these latter elements as well as titanium and cobalt [1]).

In the research on these new kinds of steel, which were sold under the name "Ticonal", not only was the composition varied in many ways, but also the nature of the heat treatment. The results were judged in each case by recording the magnetization curve and especially the remanence B_r and the coercive force H_c (see *fig. 1*). It was found that coercive forces of more than 1 000 oersted could be obtained. The remanence is usually found to decrease with increasing coercive force, so that the product of coercive force and remanence, which to a certain point may be considered as a measure of the magnetic power of the steel, cannot be increased above a certain limit of about 6×10^6 gauss-oersted.

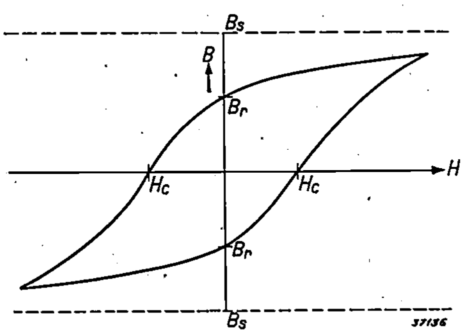


Fig. 1. Hysteresis loop of a magnet steel. B_r remanence, H_c coercive force.

Actually the quality of a magnet steel cannot be determined exactly by the product of remanence and coercive force, but by the maximum value $(BH)_{\max}$ of the product of induction and internal field strength, which may occur when the material

[1]) On the fundamental difference between the magnetic hardening phenomenon in the case of these alloys and that of the earlier known kinds of magnet steel see J. L. Snoek, Philips techn. Rev. 2, 233, 1937.

is placed in a demagnetizing field [2]). If $B_r H_c$ has a value of 6×10^6 gauss-oersted, $(BH)_{\max}$ has a value of approx. $2{\cdot}10^6$ gauss-oersted. The quotient $(BH)_{\max}: B_r H_c$ is indicated by the term convexity factor. For an exactly rectangular demagnetization curve (*fig. 2a*) the convexity factor would be equal to 1; for a straight line (fig. 2*b*) one finds 0.25. The

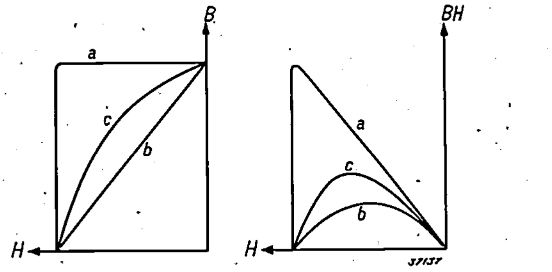


Fig. 2. *a* and *b*: Limiting cases of demagnetization curves, *c*: technical magnetization curve. Left-hand diagram *B* as a function of *H*; right-hand diagram *BH* as a function of *H*. The "convexity factor" $(BH)_{\max}: B_rH_c$ may vary between 1 and 0.25 (curve *a* and *b*, respectively); in the case of the technical curve given *c* it has a value of 0.4.

actual magnitude of the convexity factor usually lies between these two extreme values, and for the "Ticonal" type of steels it has a value of about 0.4.

On the basis of extensive attempts to reach the best possible results by variation of the composition and treatment of the alloys in question, it appeared as if the value of $(BH)_{\max}$ could not in practice be raised above $2{\cdot}2 \times 10^6$ gauss-oersted. In 1938, however, in the Philips Laboratory this value was suddenly considerably exceeded by subjecting certain new magnet steel alloys [3]) to a heat treatment in a magnetic field and then annealing them in the usual way [4]). It was found possible in this way to increase $(BH)_{\max}$ to the unequalled value of $5{\cdot}2 \times 10^6$ gauss-oersted, so that it may be stated that the quality of the steel has improved by a factor of more than 2. [5])

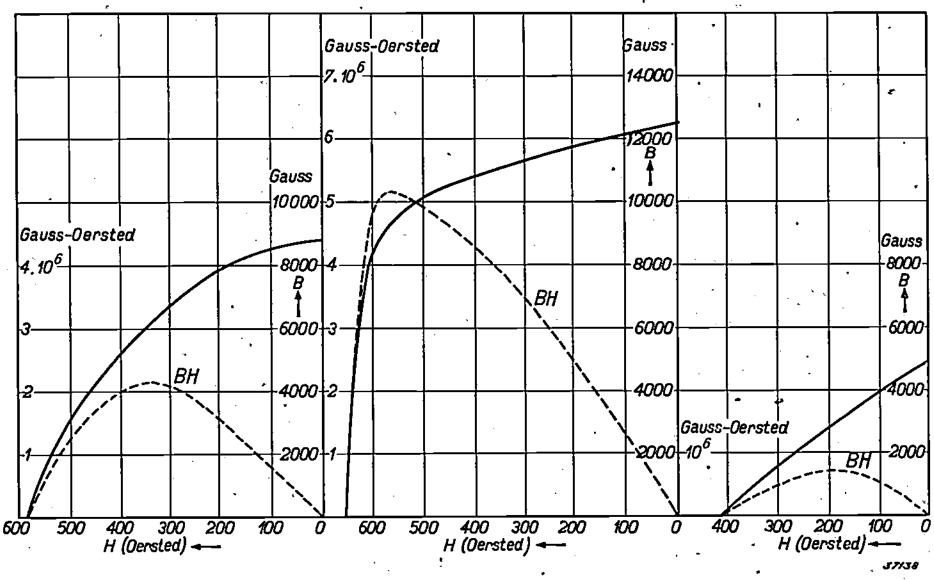
A large part of this improvement may be ascribed to an increase in the convexity factor. As may be seen in *fig. 3* the demagnetization curve of the new kinds of steel takes on a more or less rectangular shape as a result of the magnetic treatment, which means that the convexity factor begins to approach 1. The highest convexity factor measured for magnetically hardened types of steels is 0.76, and

[2]) See A. Th. van Urk, Philips techn. Rev. **5**, 29, 1940.

[3]) The alloys contain the same metal as the "Ticonal" steels previously developed, but in different proportions. A special requirement is a relatively high content of cobalt, which, in connection with the high price of this metal, at first seemed to be a disadvantage.

[4]) The application of a magnetic field during this annealing was found to be quite superfluous.

[5]) A preliminary statement on this subject was made by G. Holst before the Ned. Natuur- en Geneesk. Congress, April 11th 1939; see de Ingenieur **54**, A 199, 1939. From an article by D. A. Oliver and J. W. Shedden, Nature London **30**, 7, 1938, it appears that these investigators have carried out tests of the action of a magnetic field during the heat treatment of a so-called "Alnico" steel. Only a small increase in the remanence and the value of $(BH)_{\max}$ was observed.



a *b* *c*

Fig. 3. Demagnetization curve of a magnet steel, consisting of 51.5% iron, 8.5% aluminium, 14% nickel, 23% cobalt and 3% copper. *a*) Optimum heat treatment without magnetic field, *b*) heat treatment in a magnetic field with the direction in which the *B-H* curve is measured. In this case a value of $(BH)_{\max}$ of 5.2×10^6 gauss-oersted is obtained. *c*) The same steel as curve *b*), measured in a direction perpendicular to that of the magnetic field applied during the heat treatment. Full lines, *B* as function of *H*, broken lines, *BH* as a function of *H*.

therefore nearly twice as large as that which occurs for the same types of steel without magnetic treatment.

The remarkable character of the *B-H* curve may be explained by assuming that the elementary magnets possess a preferred orientation in the direction of the magnetic field applied during the heat treatment. In agreement with this explanation is the fact that a pronouncedly flat demagnetization curve is measured in the direction perpendicular to the field applied during the heat treatment, as is shown in *fig, 3c*. The convexity factor for this direction of magnetization only has a value of 0.32, while the convexity factor in the direction of the preferred orientation amounts to 0.66.

When curves *3b* and *3c* are compared with the curves drawn in fig. *3a* for the same material upon heat treatment without a magnetic field (convexity factor 0.43), it is seen that the field applied during the heat treatment (of about 3000 oersted) considerably increases B_r and H_c and particularly $(BH)_{max}$ in the direction parallel to it, while in the direction perpendicular to the field all three quantities are considerably decreased.

Since the beginning of 1939 the new method has been employed in manufacture on a large scale. The necessary magnetic field is generated with permanent magnets, a fact which has been made possible by the use of the new kinds of magnet steel themselves. The product on the market under the name "Ticonal" 3·8 possesses a value of $(BH)_{max}$ of 3·8 to 4 $\times$ 10^6 gauss-oersted. At the same time a remanence of more than 12 000 gauss is reached. For the sake of comparison it may be mentioned that the platinum-cobalt steel which, because of its expensiveness, could not be used for technical purposes, and which, before the development of the magnetically hardened steel, was the best magnet steel known, possessed a $(BH)_{max}$ value of 3·4 $\times$ 10^6 gauss-oersted with a remanence of only 4 000 gauss.

The results of the introduction of the new kinds of magnet steel into technology can as yet hardly be realized. It is, however, clear that not only as to new technical possibilities [6]), but also as to the price of the new magnet steel products, the prospects are very favourable. In technical respects to high remanence is a good quality in addition the the high $(BH)_{max}$ value. The high remanence makes possible constructions which exhibit a relatively small spreading of the magnetic flux, so that full advantage can be taken of the high magnetic energy per unit volume of the steel for the effective field in the air gap of the apparatus to be constructed. Since by this means a given technical problem can be solved with an unusually small amount of magnet steel, the cost price of the new magnet steel products is quite favourable. Some of the raw materials, such as cobalt and nickel, are indeed fairly expensive. If, however, the price of magnetically hardened "Ticonal" steel is calculated per unit of magnetic energy, it is found that, thanks to the high magnetic energy per unit volume, the price is lower than that of any other kind of magnet steel which has been developed in the last 20 years. Summarizing, we may therefore conclude that the new kind of steel is certainly predestined to replace in many cases the types of magnet steel used until now.

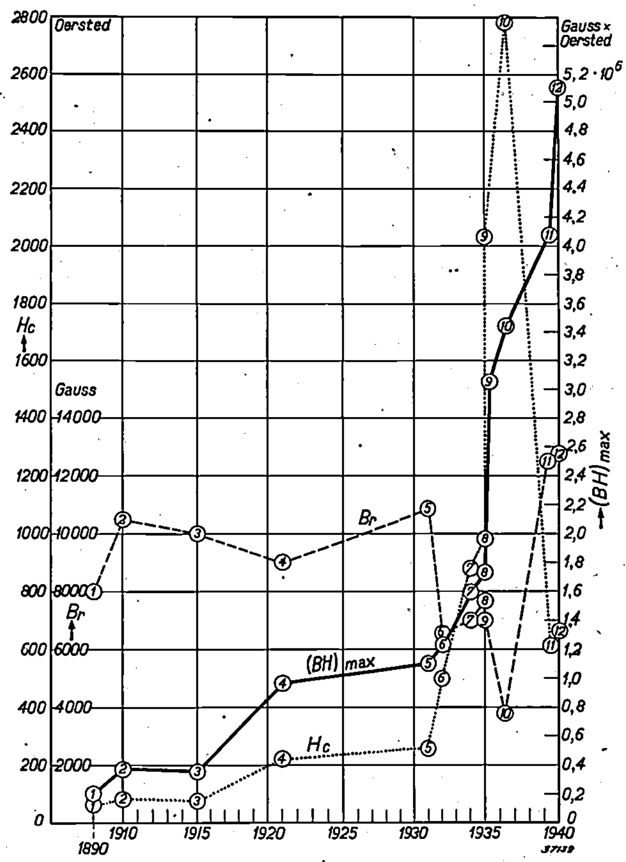


Fig. 4. The development of steels for permanent magnets during 50 years. Full lines $(BH)_{max}$, broken line remanence, dotted line coercive force. The section from 1910 to 1938 is borrowed from E. Houdremont, Stahl und Eisen, 59, 37, 1939.

1 carbon steel
2 tungsten steel
3 chromium steel
4 35% cobalt steel
5 Co-Mo steel (Köster)
6 Ni-Al steel (Mishima)
7 Ni-Co-Ti steel (Honda)
8 "Ticonal" 2 and 2a
9 Fe-Pt steel
10 Co-Pt steel
11 "Ticonal" 3.8
12 "Ticonal" 5.2

[6]) See for instance for the lifting power of the new steel, Philips techn. Rev. 5, 195, 1940.

[7]) See the article referred to in footnote [2]) page 32.

A clear picture of the development which has taken place in the last 50 years in the field of magnet steels is given in the diagram reproduced in *fig. 4*. Particular attention should be paid to the thick line which indicated the increase in magnetic power. *Fig. 5* shows the result of this progress on the construction of loud speaker magnet systems. The total weight of the systems shown (magnet steel plus coupling pieces), all of which induce the same field in a given air gap, could be reduced in steps from 1314 g to 296 g by improvement in the magnet steel; at the same time the weight of the magnet steel itself has fallen from 580 g to 74 g. The most important data about the magnets shown will be found in the following table. It is clear that the decrease in weight and volume of the magnet mean an important economy, not only for the loud speaker itself, but also for the construction of the whole radio set.

Kind of steel	$(BH)_{max}$	Weight of magnet ring	Weight of magnet system
	gauss-oersted	g	g
Cobalt steel (15% Co)	$0.6 \cdot 10^6$	580	1314
"Ticonal" 1	$1.2 \cdot 10^6$	325	703
"Ticonal" 2	$1.8 \cdot 10^6$	235	545
"Ticonal" 3.8	$4 \cdot 10^6$	104	326
"Ticonal" 5.2	$5.2 \cdot 10^6$	74	296

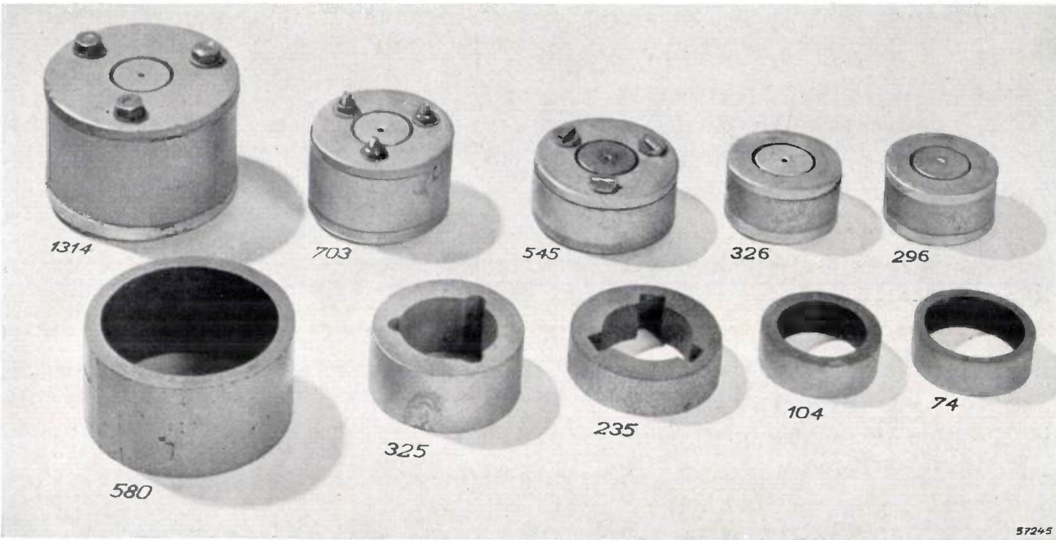


Fig. 5. A series of loud-speaker magnets which have the same field in a given air gap. The steadily diminishing size of the magnets and of the separately shown magnet steel rings gives an idea of the technical progress achieved in the course of years. The numbers indicate the weights in grams.

A UNIVERSAL APPARATUS FOR X-RAY DIAGNOSIS

by H. A. G. HAZEU and J. M. LEDEBOER.

621.386.1: 616-073

An X-ray apparatus is described which can be used for fluoroscopy and photography of any object occurring in medical diagnosis. One of the principles of its construction was that in order to obtain the best possible exposures the focus of the X-ray tube must always be loaded up to the permissible limit. This led to the result that the regulation of the tube current must not be left to the user, but must take place automatically when the voltage and exposure time have been chosen. The realisation of this principle and its influence on the construction of the rest of the apparatus is discussed in detail. In conclusion the connection of different X-ray tubes with the apparatus is dealt with, as well as its operation, which is reduced to the very simplest manipulations.

In medical X-ray examinations two different methods are used: fluoroscopy and photography. In the first case the X-rays passing through the part of the body to be examined are allowed to fall upon a fluorescent screen, and the resulting shadow picture is examined directly by the doctor. In the second case the fluorescent screen is replaced by a photographic plate or film, against both sides of which a fluorescent foil is pressed. In general fluoroscopy is used for orientation or for preliminary diagnosis, while afterwards the photograph is used for making the final diagnosis and for securing an objective documentation.

According to the part of the body to be examined and the stoutness of the patient, fluorescopy or photography must be carried out in different ways in order to obtain the best possible X-ray pictures. While in certain establishments, such as lung sanatoria, it is always the same objects (the lungs in this case) which are examined, and the X-ray installation can therefore be adapted exclusively to this special purpose, in other cases, such as non-specialized institutes, it is desirable to have an X-ray apparatus with which all kinds of very divergent examinations can be carried out. In this article we shall discuss such an installation, the Philips Medio-D apparatus, type 11 455, where the emphasis is laid on the requirement that its universal utility may not be at the expense of the quality of the X-ray picture obtained in each separate case.

Specification of the requirements

Among the variables which can be controlled in making X-ray exposures, the following are of direct importance in designing an apparatus for X-ray diagnosis: the voltage on the X-ray tube, the tube current, the size of the focus which emits the X-rays, and the exposure time (loading time of the tube). In *table I* the way in which these variables can effectively be chosen is given for several of the most commonly occurring objects. Although it would lie outside the scope of this article to give detailed reasons for the choice indicated, a brief explanation of the values given is necessary for a better understanding of the following.

Table I

Exposure technique for X-ray photographs of different parts of the body. The choice of the quantities here given is closely connected with the choice of the distance of focus to film, the kind of fluorescent foil and presence or absence of a so-called Bucky raster. It is assumed that for each of the techniques indicated these latter factors are also fixed. For the sake of simplicity, however, they have here been omitted, since they play no further part in the construction of the X-ray apparatus. The values hold for a tube with stationary anode. When a rotating anode is used, which may be loaded much more heavily (see below), considerably larger currents and shorter exposure times can be used.

Object	Voltage kV_{max}	Current mA	Diameter focus mm	Exposure time sec
Lung	55	165	3,1	$^1/_{10}$
Stomach	80*)	140	3,1	$^1/_8$
Shoulder	55	35	1,7	2
Skull	80	30	1,7	2
Lumbar (transverse)	100	25	1,7	4

*) By the administration of a paste containing barium, which is strongly absorbent, as a means of obtaining contrast, the voltage can here be given this high value.

The choice of the quantities mentioned is determined by a compromise between density, contrast and definition of the exposure obtained. Increasing tube voltage decreases the contrast because of the increasing hardness of the X-rays excited [1]), it gives, however, a greater intensity on the film (greater density), since the energy converted in the tube, the efficiency of the excitation of the radiation and the penetration of the rays increase with the voltage. With increasing tube current also

[1]) See for instance Philips techn. Rev. 2, 317, 1937. For a more detailed treatment of what is presented here in a very much simplified form see for example Philips techn. Rev. 5, 258, 1940 and the literature there cited.

a greater density is obtained because of the increase of the X-ray intensity with the energy converted in the tube. This latter process is, however, limited by the heating of the anticathode: the material of the anticathode can only stand a limited load per cm^2 of the focus (specific focal loading) without melting. When this highest permissible value has been reached, the X-ray intensity can only be further increased by taking a larger focus. At the same time, however, the definition of the picture decreases due to the larger half-shadow width. Finally the density also increases with the exposure time; but when it is a question of moving objects, such as lungs, stomach, etc., this is accompanied by a greater lack of definition.

The values of table I represent (for a given allowable loading of the focus) practically the most favourable compromise which is to be found under these opposing influences in the various cases. The way in which this compromise is arrived at for the properties of the objects to be photographed is now fairly clear: upon photography of moving objects such as lungs and stomach the exposure must be very short, so that a large focus is needed; in skull photography and the like, the exposure may be for several seconds, and therefore greater definition can be expected by the use of a smaller focus; with very absorbent objects (with large contrasts), such as bones, the voltage is made high, with weakly absorbing tissues, such as the lungs, low voltages are demanded, etc. One point is, however, fixed in all these cases: in order to reach an optimum result, the focus must be loaded up to its limit. The permissible loading (product of current and voltage) still depends upon the time of loading of the tube (exposure time): with short loading the load may be heavier than with a long time of loading. Therefore if two of the three quantities, current, voltage and time are chosen, for instance the voltage and the time, then the third, the current, may no longer be considered as an independent variable, but it is fixed by a relation which differs for every X-ray tube, an example of which is given in *fig. 1*.

If one compares the optimum attainable quality of the photograph for different X-ray tubes, then under otherwise similar conditions that tube will be the best in which the permissible specific focal loading [2]) is the highest.

The foregoing has referred to photography. As to fluoroscopy, one must always accept a poorer quality of the X-ray picture, since the loading time is here very much greater (several minutes), and therefore the permissible loading of the tube much smaller [3]). Moreover, the total X-ray dosage may not exceed a certain value for the safety of the patient. The best compromise attainable with the limited intensity between contrast and clearness of the picture can be found by the doctor himself during the examination by varying the voltage somewhat, to for instance 10 kV higher or lower than the value used for the photograph.

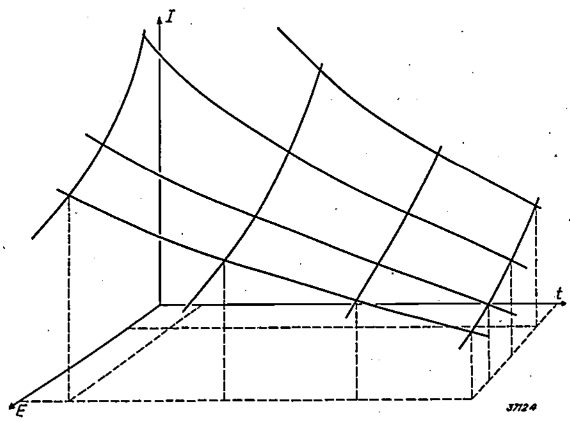


Fig. 1. Relation between tube current I, tube voltage E and loading time t, when the focus is loaded up to the highest permissible temperature. The surface drawn is for an X-ray tube with rotating tungsten anticathode and for loading with so-called commutated alternating current (4 valve connection).

We may now make several stipulations which must be met in the construction of the universal apparatus for diagnosis. The tube voltage must be variable between several hundredths of a second and several seconds; the current must always be such that the focus is fully loaded, which means a variation between 25 and 500 mA, according to the tube used and the object to be examined. It is obvious that for obtaining comparable photographs the values chosen must be easily reproducible.

In addition to these requirements, which are connected with the possibilities of variation of the exposure technique, there are several others connected with the placing of the patient and the method followed in the examination, which exert their influence on the construction of the apparatus. The X-ray tube and the film holder must be hung on a standard which permits a rapid and accurate adjustment of focus and film with respect to the patient. For the examination of different organs different standards will sometimes have to be used,

[2]) By specific loading is here meant the loading per cm^2 of the "apparent" focus; see in this connection Philips techn. Rev. 3, 262, 1938.

[3]) The loading limit here is not set by the danger of melting of the anticathode at the focus, but by the danger to which the glass parts of the tube are exposed by too great general heating of the anticathode.

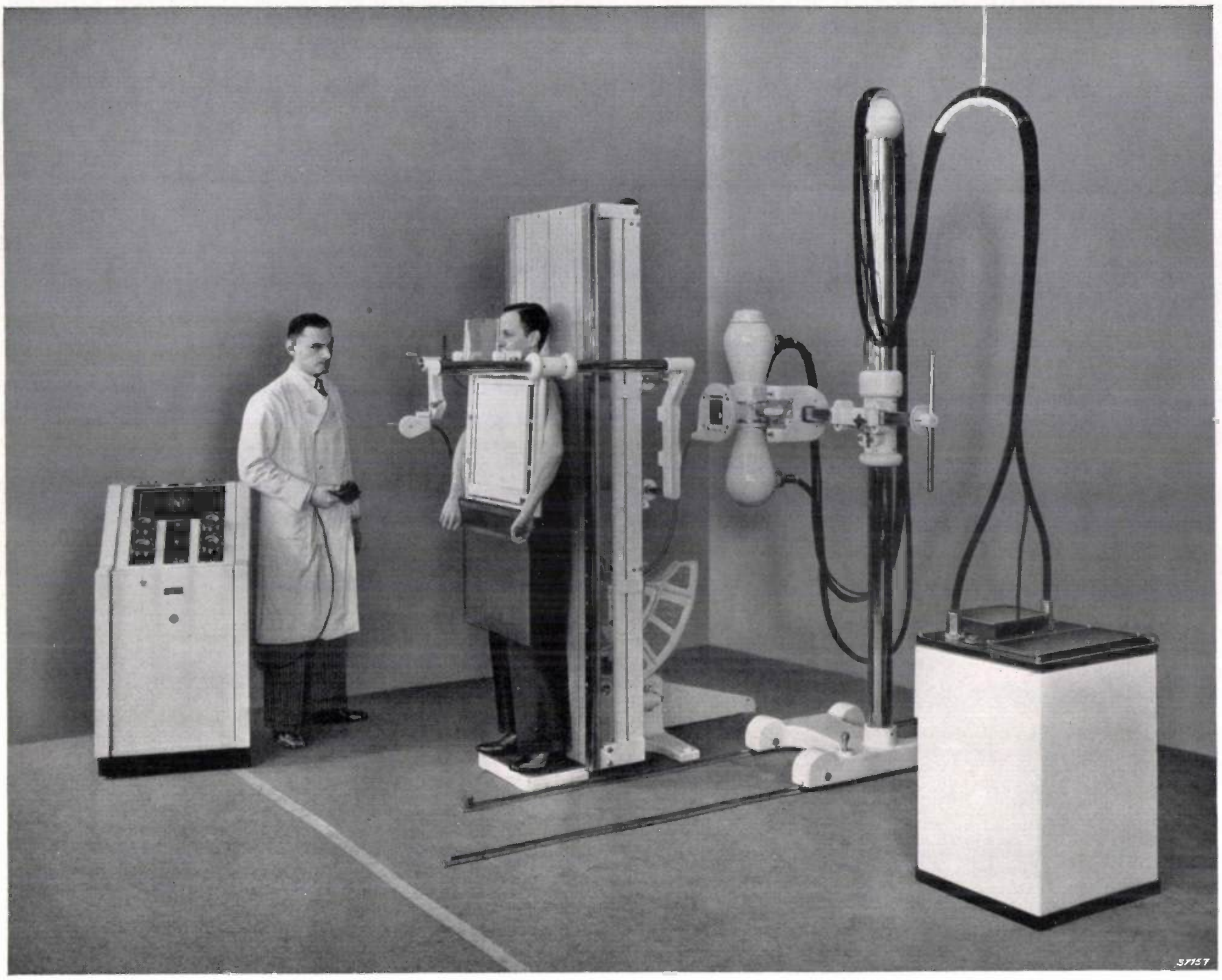


Fig. 2. View of the whole X-ray apparatus at the moment when a lung exposure is being made. On the right the iron tank filled with oil containing the high-voltage generator; on the left the operating desk. In the middle the standard which bears the fluoroscope screen or the film holder, behind that the X-ray tube hung on a movable column.

each with its own tube. The switching arrangements of the tube and any auxiliary apparatus, as well as the adjustment of the exposure quantities must be so simplified that the attention of the doctor is not thereby distracted. In certain cases, such as stomach examinations, it is also desirable by means of fluoroscopy to be able to determine the moment at which the object has taken on the position or shape which is to be fixed on the photograph. The apparatus must then be so arranged that the exposure can take place as quickly as possible after the fluoroscopy. In connection with this the variation of the voltage in fluoroscopy may not affect the previously made adjustment for the exposure, etc.

We shall now show how these different requirements are met in the apparatus to be described.

General description of the apparatus

The apparatus consists chiefly of the tube with standard, a high-voltage generator for supplying the tube, a timing switch for switching the tube load on and off, and different regulatory devices which are combined in an operating desk. In *fig. 2* a photograph of the complete installation is given, while *fig. 3* shows the general plan of the connections.

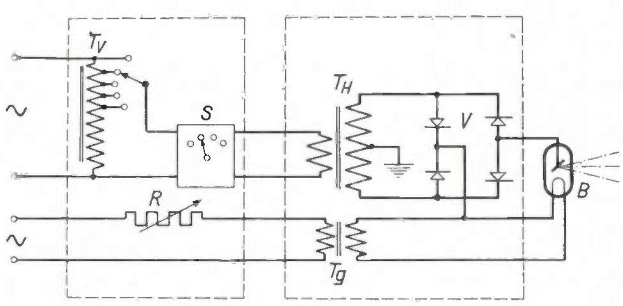


Fig. 3. Connections of the apparatus (very much simplified). T_H high-voltage transformer, V rectifier valves, B X-ray tube, S timing switch. T_v auto-transformer for regulating the tube voltage, T_g heating-current transformer, R resistance for varying the tube current.

The high-voltage generator contains a transformer connected to the local mains, which furnishes the required high-voltage of $E_{max} = 100$ kV. This voltage is rectified with four valves in the familiar Grätz connection, so that a pulsating D.C. voltage with a peak value of 100 kV is obtained on the tube (*fig. 4*). The variation of the tube voltage is obtained by changing the primary voltage of the high-voltage transformer with the help of taps on an intermediate autotransformer. The heating current of the tube is provided by a separate heating-current transformer; by means of an adjustable resistance in series with the primary winding the heating current, and with it the electron emission of the cathode, and thus the current through the X-ray tube, can be varied. The desired exposure time is finally obtained with the help of the above-mentioned timing switch which will be described below, and with which, it may be mentioned, the loading time of the tube can be adjusted between $^1/_{50}$ sec and 8 sec in steps.

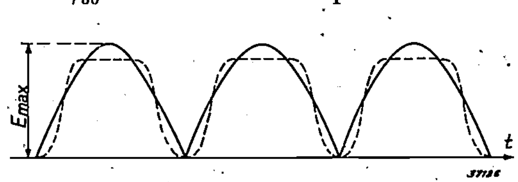


Fig. 4. Variation of the tube voltage with the time. The broken line gives the variation of the current which is in phase with the voltage. In each half period the current reaches a saturation value at a certain voltage. This current is given by the cathode emission.

The automatic character of the adjustment

In the above we have seen that when optimum results are desired the tube current must not be considered as an independent variable, but that with voltage and loading time determined it is fixed for a given tube by a relation as in fig. 1. In the case of the X-ray apparatus developed by Philips the obvious conclusion has been drawn that in photography the current should not be regulated by the operator, but upon regulation of the voltage (which mainly determines the contrast) and the loading time (which determines the lack of sharpness due to motion) it must automatically take on the corresponding highest permissible value.

Such an automatism can in principle be realized in a very simple way, see for example *fig. 5*.

To the axis of the voltage regulator is coupled a "pre-selector", *i.e.* an arm which can connect one end of the primary winding of the heating-current transformer successively to as many "main selectors" as there are voltage steps. The arms of the main selectors are all coupled with the axis of the regulator of the loading time, and every main selector can continue the above-mentioned connection with as many taps of the heating-current resistance as there are time steps. If the apparatus has m voltage and n time steps, $m \times n$ taps are made on the heating-current resistance; these taps are so arranged when the whole installation is adjusted that for each of the $m \times n$ possible combinations of voltage and time values the proper current value is obtained.

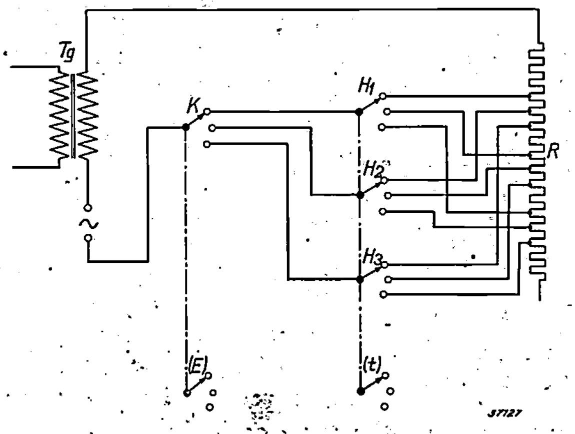


Fig. 5. Diagram showing the principle of the automatic adjustment of the tube current. T_g heating-current transformer, R resistance with as many taps as there are possible voltage-time combinations. K pre-selector coupled with the voltage regulator (E). $H_{1, 2, 3} \ldots$ main selectors coupled with the time regulator (t).

The voltage must be adjustable in steps of about 2.5 kV; for the whole range of variation from 50 to 100 kV, therefore, about 20 voltage steps are needed. The same number of steps is needed for the exposure time, so that there would be no fewer than 400 taps on the heating-current resistance. Fortunately it is found in practice that a considerably smaller number is sufficient. The *E-I-t* surface of fig. 1 can be approximated by a stepped surface like that shown in *fig. 6*. All the voltage steps are here divided into three groups, and the

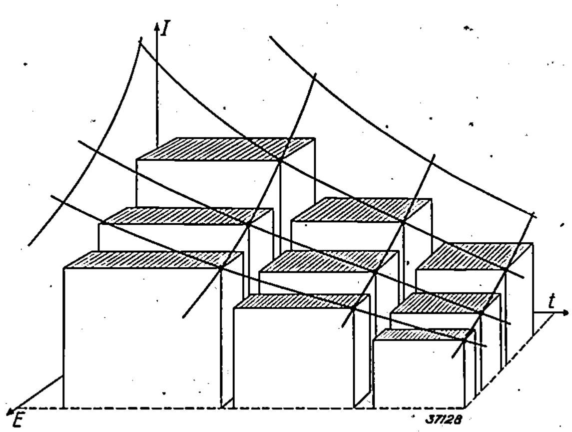


Fig. 6. Approximation to the curved surface of fig. 1 by a stepped surface. On the cross-hatched plane surfaces of the subdivisions the current is constant.

same is done with the time steps. For each voltage group in the automatic arrangement there is one main selector which selects one tap for each time group [4]). One current value is thus selected for each of the nine group combinations. The highest permissible loading is now only attained for the highest voltage and the longest time of each group combination. At lower voltages and with shorter times in the same group the loading may remain as much as 30 per cent below the permissible value; no important difference in the quality of the exposure compared with the optimum results from this fact, however.

For fluoroscopy a variable resistance operated by hand is connected in the heating-current circuit. By means of a series resistance the tube current is prevented from ever increasing above the permissible value during the fluoroscope examination.

The voltage loss

Until now we have represented the case as if the voltage on the tube could be adjusted quite independently of the tube current. Actually account must be taken of the fact that upon flow of current through the tube a certain voltage loss occurs which increases with the current. This loss, which in certain apparatus may amount to 20 kV, is caused by the resistances of the high-voltage transformer, the switch arrangement, the connection lines to the generator, etc. and by the voltage drop in the rectifiers. It would of course be possible to compensate for the voltage loss simply by a correspondingly higher setting of the transformer voltage. This has, however, disadvantages connected with the automatism here employed. If for instance an exposure time of $^1/_5$ sec is desired with a tube voltage of 60 kV, the automatic arrangement provides a tube current of 260 mA for a given tube. Suppose that the voltage loss here amounts to 20 kV, the transformer voltage must then be 80 kV (peak values are meant in every case). If the operator now raises the voltage one step, *i.e.* to 82·5 kV, the following voltage group is reached, where with $^1/_5$ sec a current of only 210 mA is provided; the voltage loss here is considerably less, 16 kV for instance, the tube voltage is therefore 82·5 — 16 = 66·5 kV. The whole voltage range between 60 and 66·5 kV is thus traversed in one step without the possibility of a finer adjustment.

In order to avoid this difficulty attempts were made to limit the voltage loss as much as possible in the whole apparatus. To this end in the first place a heavy high-voltage transformer with only slight resistance is used. Likewise for the voltage regulation an autotransformer with thick windings is taken and a rotating switch with broad contacts and thick connections. Furthermore for the rectifier connections hot-cathode valves with gas filling are used which have a voltage drop in the transition direction of only about 50 volts [5]). A not inconsiderable part of the remaining voltage loss is caused by the resistance of the local mains themselves, since the apparatus can take up for short times energies up to 15 kW. Since this mains resistance may vary very much in different installations, while it is nevertheless desirable that the same voltage losses and thus the same exposure values should be obtained, an additional variable resistance in series with the mains connection is included, which supplements the local mains resistance to give the same value for all installations. In *fig.* 7, which shows several other details of the primary circuit, this is illustrated more fully. The voltage loss in the mains at 260 mA now always amounts to 4 kV; added to this are 3 kV loss in the high-voltage transformer, 0.1 kV in the valves, etc., giving a total loss of about 10 kV.

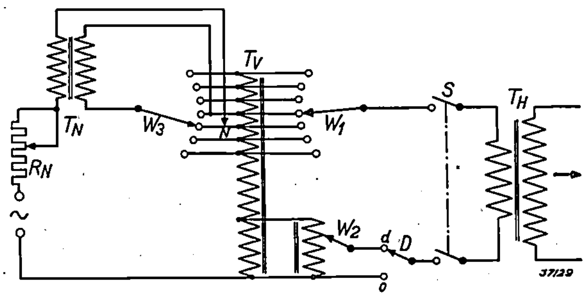


Fig. 7. The primary voltage of the high-voltage transformer T_H is regulated by selecting different taps of the auto-transformer T_v with the contact arm W_1. For fluoroscopy switch D is placed on d (o is for photography), and the voltage can be varied an additional ± 10 kV with the help of W_2. By suitable choice of the connection N the apparatus can be used at different nominal mains voltages (between 150 and 440 V). In order always to obtain the same high-tension values with a given setting of W_1 and W_2 upon variation in the mains voltage, the input voltage of the auto-transformer can be varied slightly further with the help of the auxiliairy transformer T_N and the switch arm W_0. S main switch. R_N variable resistance with which the voltage loss can be made equal for all installations.

Reproducibility of the adjustment

The type of automatic arrangement described affects the construction of the whole apparatus, not

---

4) Actually the automatic arrangement in the apparatus described is slightly different; for the sake of simplicity, however, this simple description may be considered as a basis.

5) These high-voltage valves have already been described in Philips techn. Rev. 1, 8, 1936.

only because of the requirement of a low voltage loss, but also because of the high requirements made of the reproducibility. Since the tube current can here be varied by changing the heating current, it is necessary in the first place that the electron emission of the hot cathode should be absolutely constant. When the emission decreases somewhat during the lifetime of the tube, the quality of the photographs suffers. In the regularly occurring checks therefore the taps on the heating-current resistance must be regulated anew and adapted to the changed emission properties. More dangerous than this gradual drift would be the variations of the heating voltage and of the resistances in the heating-current circuit occurring during operation, since the emission depends very closely upon the heating current and since instead of a decrease in the quality of the picture, an overloading of the tube could occur. The heating voltage is kept constant, *i.e.* made independent of mains voltage fluctuations (and of any mains voltage fall due to the high current during exposure) by connecting a "stabilizer" in front of the heating-current transformer [6]). The resistances in the heating-current circuit must above all be protected from becoming too hot, which would cause the specific conductivity to vary. This is especially true for the heating-current transformer which is housed together with the high-voltage transformer and the rectifiers in an oil tank, and which therefore might become too hot due to the proximity of the latter. Thanks, however, to the above-mentioned heavy build of the transformer and to the use of gas-filled valves in which oxide cathodes can be used and which therefore give the necessary emission with about 10 W heating-current energy, the heating up of the whole generator container is reduced to a harmlessly low level.

The timing switch

Like current and voltage, the loading time must be very accurately reproducible, to within a few per cent, for instance, in order to load the focus correctly. This is no simple condition when it is kept in mind that it is here a question of the switching of powers of up to 15 kW (according to the X-ray tube employed) during times which must be able to be adjusted between several hundredths of a second and several seconds.

The circuit diagram of the timing switch developed for this purpose is given in *fig. 8b*. In order to make its action clear, let us consider first the very much simplified diagram of fig. 8*a*. The switching of the tube voltage (main switch S) takes place by means of a relay Sp which is excited by a transformer T_1 *via* the relay valve L_1. Such a relay valve (gas-filled triode) can only ignite when the grid as well as the anode have a sufficiently high voltage with respect to the cathode, and only be extinguished when the anode voltage becomes negative (or falls below a certain value). As long as the push-button switch A is open, the grid G_1 is positive with respect to the cathode. After A is closed therefore the valve ignites as soon as the anode is positive enough; the relay receives current and closes the switch S, the loading of the X-ray tube begins. As the supply A.C. voltage passes through zero the relay valve is extinguished, while a half period later it again ignites. If we now first assume that the time of opening of the relay is so long that the currentless periods (during which the anode of L_1 is negative) are thereby bridged over, then the high-voltage circuit will always remain closed. Simultaneously with the closing of the main switch S, however, the relay has opened the auxiliary switch F, so that the battery B begins to charge the condenser C *via* the resistance R_1. The grid G_1 connected to the upper condenser plate is hereby gradually made less positive, and after a certain interval of time, which can be regulated by adjustment of R_1, it becomes negative with respect to the cathode. Now the reignition of the relay valve after the next negative period of the anode voltage is no longer possible, the relay Sp falls out, the loading of the X-ray tube is ended. (Special measures have been taken, so that the swtich does not "repeat", *i.e.* so that it does not immediately switch on again directly after the restoration of the initial state).

By the variation of R_1 (and if necessary of C) the loading time can be regulated within wide limits. The reproducibility would, however, by no means be satisfactory. According as the moment when A is switched on falls at the beginning or end of the positive period of the alternating current, the loading time may already vary by $^1/_{100}$ sec (a half period). Still worse are the differences which may occur by the more or less rapid falling out of the relay, which, according to the above, must work with a relative time lag. These disadvantages are avoided as follows in the actual construction (fig. 8*b*). For the excitation of the relay, a second relay valve L_2 is introduced, whose grid is normally kept negative by the D.C. voltage source B_1. If, however, the valve L_1 is extinguished, a voltage surge is caused in the secondary winding Sp_2 of the relay Sp, which just makes the grid of L_2 pos-

---

6) See Philips techn. Rev. 2, 276, 1937.

itive at the moment when the anode voltage of L_2 is positive, so that L_2 now ignites and transmits current for half a period. The valve L_1 thus always "follows" the valve L_1 and supplements the direct current impulses produced by the latter in the intermediate half periods. It is not necessary for this reason to give the relay a long opening time for bridging over these half periods. On the contrary, by making the opening time as short as possible the fluctuation in this time is also decreased and the reproducibility increased.

The connection of different X-ray tubes

The explanation given at the beginning on the technique of exposure leads to the conclusion that for every object a definite optimum size of focus should be chosen. The size of the focus in a given tube is permanently fixed by the shape and dimensions of the cathode. Thus for every kind of object one would need a separate X-ray tube with a definite cathode. For smaller hospitals where financial reasons make this condition a handicap, a tube has been constructed which contains two separate hot cathodes which can be used at will. By this means at least two different foci are available, for instance one of 3.1 mm diameter for moving objects (lungs, stomach, etc.), and one of 1.7 mm diameter for stationary objects (skull, shoulder, etc.). A better solution is the use of an X-ray tube with rotating anode ("Rotalix" tube [7])) in which the permissible specific focus loading is a factor 6 to 10 higher than with a stationary anode, and with which therefore even with a small focus sufficient intensity for exposures of moving objects is obtained.

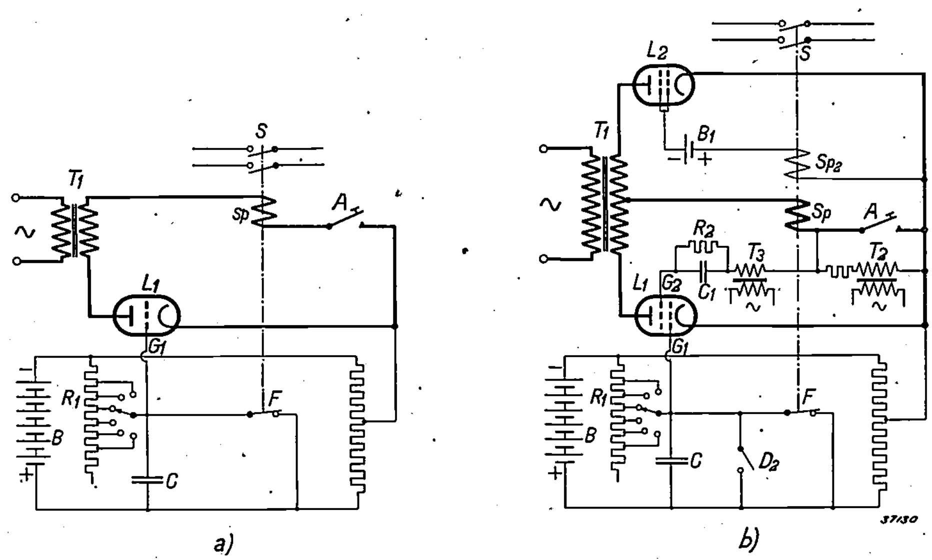


Fig. 8. Timing switch. *a*) Principle of the timing arrangement: after setting the push button switch A the relay valve L_1 ignites, the relay Sp switches the high-voltage transformer in through S, but at the same time opens by means of F the short circuit of the condenser C. The speed with which C is charged by the battery B, *i.e.* the setting of the variable resistance R_1, determines the time during which the relay valve can transmit current.

b) More complete diagram. In these connections "isochronous" switching of the relays takes place at the zero points of the voltage. The condenser C_1 in the stationary state (A open) is charged through the auxiliary transformer T_2, and, thanks to the rectifying action of the hot cathode, in such a way that the second grid G_2 of the relay valve L_1 is negative. Upon closing A therefore the relay valve cannot immediately ignite. Due to the fact, however, that the charging voltage of C_1 is now short-circuited, C_1 discharges gradually over R_2, the potential of G_2 slowly rises to cathode potential. Through the small transformer T_0 an A.C. voltage acts on G_2 which causes G_2 to become positive just at the moment at which the anode also becomes positive. At this moment therefore the valve ignites, independent of the moment when A is closed. Switching off always occurs "isochronously" of itself, since the relay valves are always extinguished, independent of the grid voltage, when the anode voltage becomes negative.

In order to obtain a short opening time the movable parts of the relay must be kept light. This would not be possible if the relay had to switch over the whole energy, but in that case heavy contacts would be necessary because of the wear. Therefore provision has been made that the switching on and off takes place "isochronously" *i.e.* always at moments when the voltage (and therefore in our case the current also) passes through zero. At the same time by this device the previously mentioned inaccuracy due to the arbitrariness of the moment of closing switch A is eliminated. The way in which the isochronous switching is ensured is explained further in the text under fig. 8.

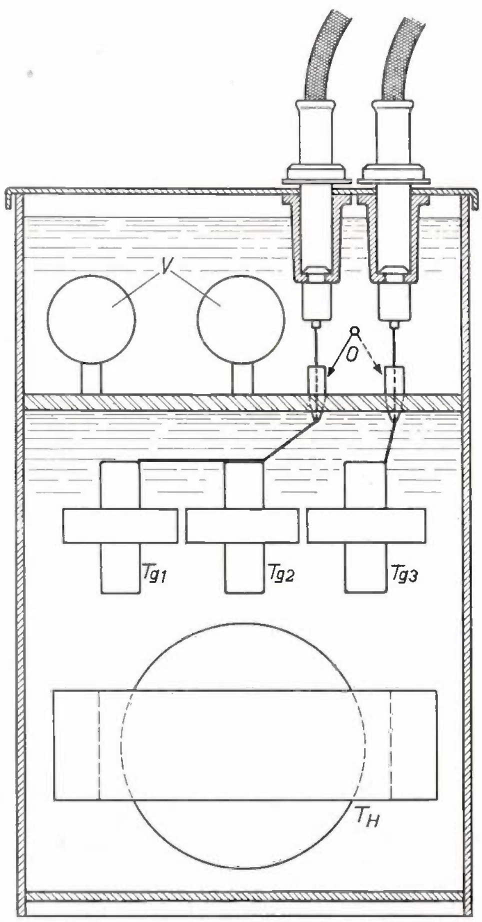
---

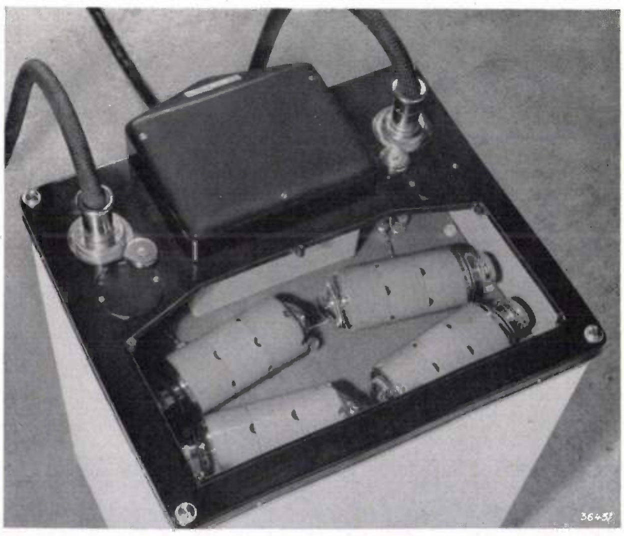
7) See Philips techn. Rev. **3**, 292, 1938.

The apparatus here described is specially designed for connection with these two tubes, and the generator is provided to two sets of cable connections, the choice between them being made with a built-in high-voltage switch. Several problems were encountered, particularly in connection with the cathode supply. The cathode, and with it the secondary winding of the heating-current transformer are under high tension, since the middle point of the tube voltage is earthed in order to have to insulate only against half the high-voltage with respect to earth. The fact that the cathode is under high tension led on the one hand to the heating-current transformer being housed in the oil tank of the generator (see above) and on the other hand it made it desirable to carry out as few switching manipulations as possible in the cathode circuit. Therefore three heating-current transformers are housed in the generator (see *fig. 9*). The secondary windings of two of them are connected to a post of one set of high-voltage connections and can supply the two cathodes of a tube with double focus. The secondary winding of the third is connected to a post of the second set of high-voltage connections. Thus upon passing over to a different tube or focus switching need only be performed in the primary circuits of the heating-current transformers.

In addition to the switching over of the high voltage and the cathode supply, care must also be taken that the automatic arrangement of the tube current regulation is correct for each tube (or each cathode). This is realized in a simple way by using a separate heating-current resistance for each tube (or cathode), the nine taps of which resistance are adjusted separately. *Fig. 10* shows how these automatic units are assembled in the operation desk. Upon connection of still other X-ray tubes with the apparatus (for which a separate high-voltage commutator is necessary) more of such units, up to a total of six, can be built in.

The commutation of the high-voltage takes place simultaneously with the commutation of the cathode, of the series resistance for the fluoroscope heating current (see above) and of the automatic arrangement, by means of a relay, all of which are operated at once by a tube selector mounted on the desk (*fig. 11*).





a *b*

Fig. 9. *a*) Assembly of the high-voltage transformer T_H, the three heating-current transformers T_g and the four rectifier valves V in an iron container filled with oil (dimensions about 80 × 50 × 50 cm). Since the valves must be accessible for control, replacement, etc., while the transformers, in whose case the insulating oil also serves as impregnating medium, must remain carefully closed, the container is divided into two parts by a horizontal partition. In fig. 9*b* a view is shown of the upper part (not yet filled with oil) in which the rectifiers are housed. In this part is also the high-voltage switch *O* for the two sets of connections.

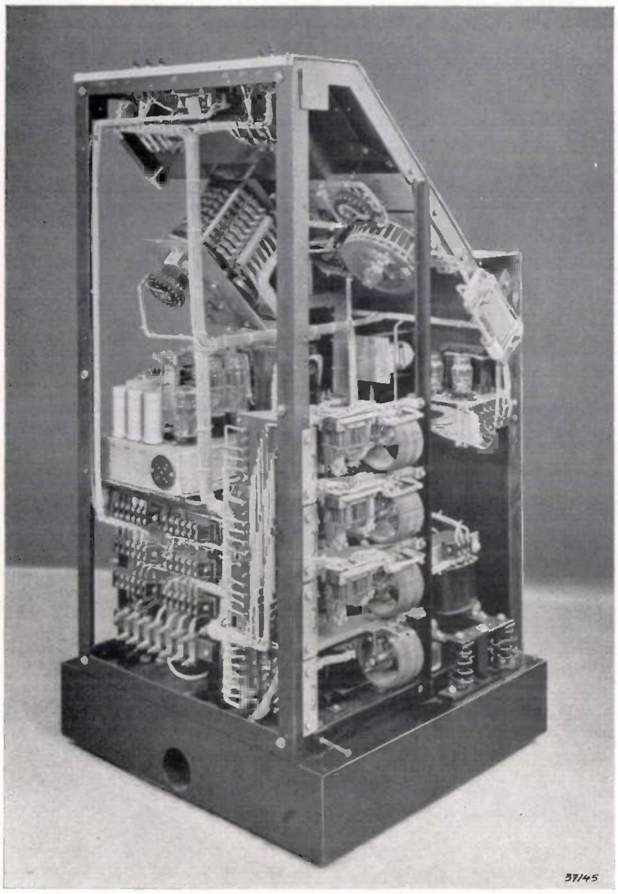


Fig. 10. View of the interior of the operating desk. To the right in front may be seen three "automat" units mounted one above the other, *i.e.* heating-current resistances with switch contacts for three different cathodes. If more tubes are connected to the apparatus a unit is added for each cathode. Under the cover are the rotating switches of the voltage regulator, etc. To the left below, the timing switch.

Using the apparatus

The apparatus is so arranged that after switching on the mains voltage everything is normally ready for fluoroscopy: the tube receives a small heating current, the separate voltage regulator (W_2 in fig. 7) is switched on, the timing switch is out of action (D_2 in fig. 8*b* closed). In the meantime, however, the desired voltage and time for an exposure can already be set, without affecting the fluoroscope image.

For photography, in addition to setting the various switches (D in fig. 7, D_2 in fig. 8*b*, measuring range of the mA and mA-sec meters K and L on the operating desk, etc.) the hot cathode must furnish the higher emission chosen, and when a "Rotalix" X-ray tube is used the anode must have reached its working speed. This requires some time, about 0.8 sec in both cases. The doctor using the apparatus need not bother with all these things, however: all switching operations are accomplished automatically by pressing on the knob of a hand switch, while a retarding relay provides that the timing switch only begins to work after 0.8 sec.

Other mechanisms also which are involved with photography can be coupled with the hand switch, for instance a moving raster for attenuating the scattered rays or a mechanism for bringing the film holder into position. This is necessary particularly in the so-called stomach series examination which was touched upon at the beginning. In this case by pressing the knob "photography" a film is also inserted in front of the fluoroscope screen by an electrically operated mechanism, so that between the fluoroscopic examination and the photographic exposure no more than the 0.8 sec mentioned elapses. In this way the required operations are reduced to the adjustment beforehand of time and voltage and the pressing of the knobs "fluoroscope" and "photograph".

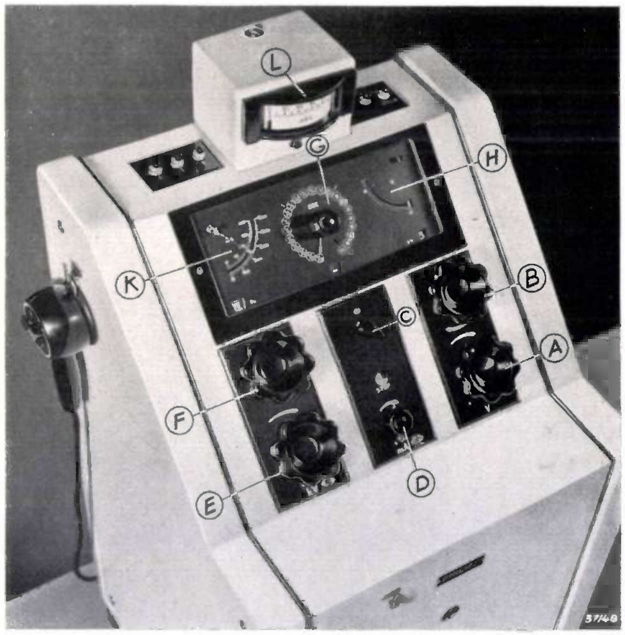
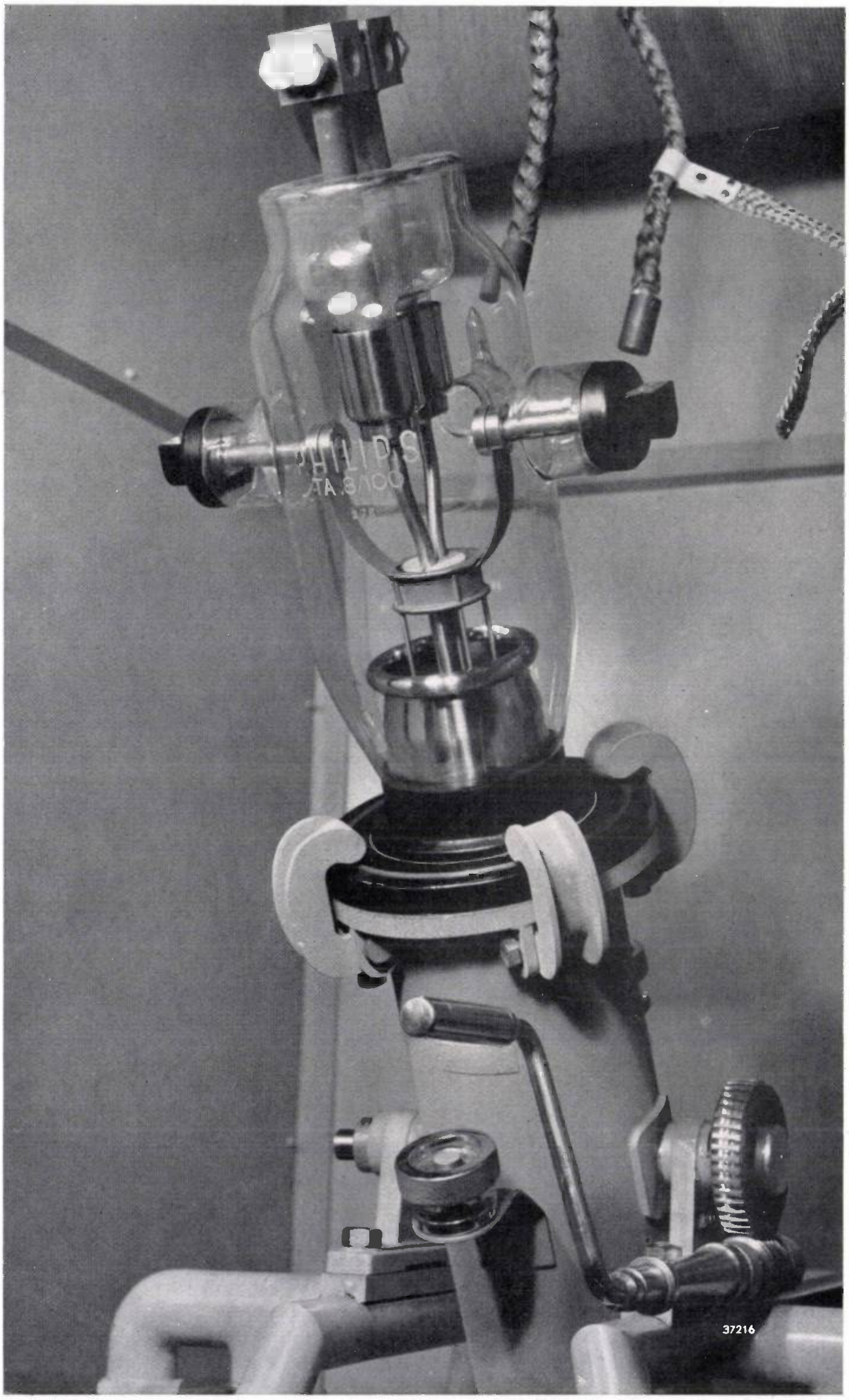


Fig. 11. The switchboard of the operating desk. By means of the so-called tube selector F, high-voltage, heating voltage, automatic arrangement, etc. are all switched over at once to the X-ray tube to be used. The other knobs serve for the regulation of the time (G), the coarse and fine regulation of the tube voltage (B and C), the regulation of voltage and current for fluoroscopy (E and D) and the correction of the mains voltage (A; checked with voltmeter H). On the left hanging on a hook, a hand switch with which the doctor standing at any desired distance from the desk can switch the apparatus on and off in fluoroscopy or photography.

THE WATERCOOLED TRANSMITTING VALVE TA 18/100



TA 18/100 is a watercooled transmitting valve, which in H.F. class C telegraphy adjustment supplies 100 kW and about 38 kW in the carrier wave, using anode modulation.
The maximum allowable anode dissipation amounts to 70 kW. Filament voltage 33 V, filament current 207 A.
Overall length without cooler 120.5 cm, with cooler 133.3 cm.
The photograph shows this valve in a special arrangement, as used in the new Netherlands broadcasting system, which facilitates easy interchangeability of valves.

THE RECORDING OF DIAGRAMS WITH THE ELECTRICAL PRESSURE INDICATOR

by P. J. HAGENDOORN and M. F. REYNST. 531.787.9

Following a previous article in this periodical which gave a detailed description of the electrical pressure indicator for internal combustion engines developed by Philips, a further study is here made of the different kinds of diagrams which can be recorded with this apparatus. The detailed construction is given of the piston-stroke recorder with which a deflection of the cathode ray proportional to the displacement of the piston is obtained. For the routine testing of large engines, in Diesel stations for instance, special devices have been developed which are also briefly discussed.

With the help of the pressure indicator GM 3 154, which was described in a previous article [1]), the variation of pressure in the cylinder of an internal combustion engine can be made visible on the screen of a cathode ray tube. The pressure variations are first converted into capacity variations of a condenser, one of whose electrodes is formed by a membrane in the wall of the cylinders (pressure recorder). The capacity variations obtained are used to modulate a carrier wave from which, after amplification and rectification, the required voltage for the vertical deflection of the fluorescent spot is obtained. As to the horizontal deflection, two methods may be used: it may be made proportional to the time or to the displacement of the piston, thus to the volume (more exactly: to the increase in the volume) of the combustion chamber. In this article we shall study the way in which these two types of diagrams and all kinds of variations of them are obtained, as well as the various auxiliary apparatus which have been developed for this purpose.

Pressure-time diagrams

A horizontal deflection of the cathode ray proportional to the time is obtained by applying to the proper set of plates of the cathode ray tube a voltage of the form given in *fig. 1*. In order to generate this sawtooth voltage a time-axis generator is used in the pressure indicator which corresponds exactly to that of the cathode ray oscillograph GM 3 156, which has recently been described in this periodical [2]).

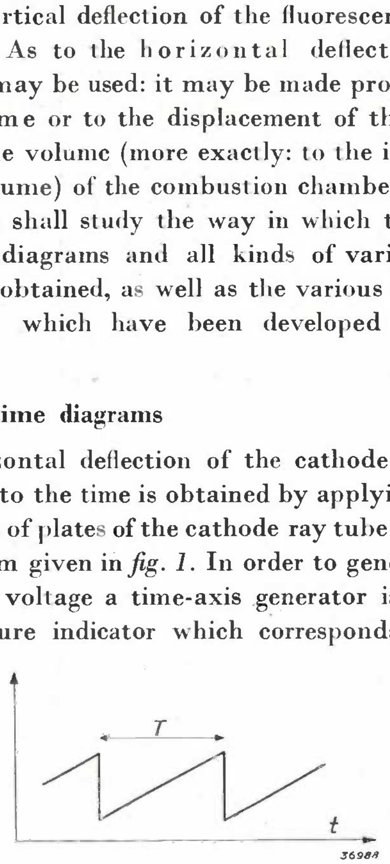


Fig. 1. Sawtooth voltage for the recording of diagrams with the time as abscissa.

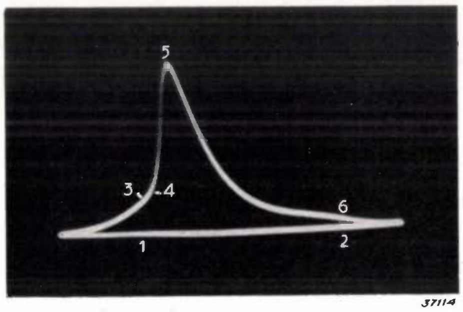


Fig. 2. Normal pressure-time diagram of a four-stroke engine recorded with the pressure indicator GM 3 154. The time base is equal to one revolution. This photograph, like those of figs. 3, 4, 5, 12, 13 and 14, has been put at our disposal by the *Bataafsche Petroleum Maatschappij*, Testing station, Delft. Fig. 12 in the article referred to [1]) is also from this testing station.

Fig. 2 shows a normal pressure-time diagram of a four-stroke engine recorded in this way. In order to obtain a stationary image on the fluorescent screen it is necessary that the period of the time base (T in fig. 1) should be exactly equal to half the fundamental period of the diagram, thus to the time necessary for one revolution around the crank-shaft. By regulation of the time-axis generator the period of the time base can be set approximately at the desired value, while exact synchronisation with the engine is realized by an extra voltage surge which, with Diesel engines, is supplied once per revolution *via* a contact disc mounted on the crank-shaft to the time-axis generator, and with petrol engines, by means of the voltage impulse of the sparking plug.

In the diagram of fig. 2 the following processes may be distinguished: the sucking in of the gas mixture into the combustion chamber (*1-2*) at a

[1]) P. J. Hagendoorn and M. F. Reynst, An electrical pressure indicator for internal combustion engines, Philips techn. Rev. 5, 348, 1940.

[2]) S. L. de Bruin and C. Dorsman, A cathode ray oscillograph for use in tool making, Philips techn. Rev. 5, 277, 1940.

pressure slightly less than one atmosphere, the compression of the air (*2-3*), at *3* the ignition and beginning of the combustion (*3-4*), the combustion (*4-5*), at which the pressure quickly reaches a peak value (*5*), the expansion of the burned and heated gases (*5-6*), the driving out of these gases (*6-1*).

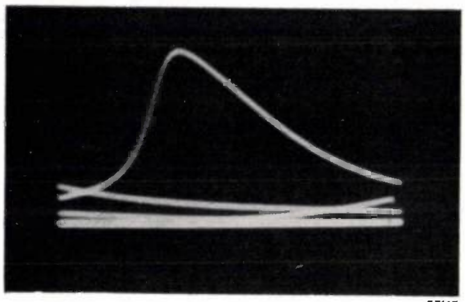


Fig. 3. The same diagram as fig. 2 recorded on a larger scale. The time base here was made equal to half a revolution of the crank-shaft.

While the sucking in, compression, expansion and driving out are relatively simple processes, more or less fixed by the dimensions of the cylinder, the valve openings, etc., the pressure variation from *3* to *5*, *i.e.* the process of combustion, depends upon many influences which are not directly controllable. Here therefore deviations may most easily occur from the normal action, and here therefore the control by means of the diagram is most important. With the time-axis generator the diagram can also be recorded on a larger scale, so that the details of the part in question are clearer. For this purpose the time base is adjusted to a length equal to a half, a third, a fourth, ... revolution. One then obtains, for example, on the fluorescent screen an image like that of *fig. 3*. The synchronization arrangement here also provides that the image is stationary.

In the case of internal combustion engines one of the most feared phenomena is so-called detonation which is manifested by the appearance of vibrations in the expansion lines (see *fig. 4*). Not only due to

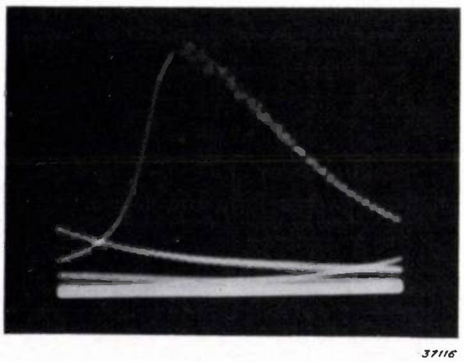


Fig. 4. Pressure-time diagrams with detonation vibrations on the expansion line.

too heavy loading, incorrect composition of the gas mixture, carbon deposit in the cylinder and similar causes, detonation may also occur due to too high compression and irregular ignition, resulting for instance from too high a temperature of the cylinder wall (insufficient cooling).

In certain cases it may be desirable to employ no synchronization. The following is an interesting example. If the time-axis generator is set as in fig. 2 but synchronization is omitted, the variation of pressure for the successive revolutions will not always appear at the same spots on the fluorescent screen but will gradually shift, see *fig. 5a*. In the

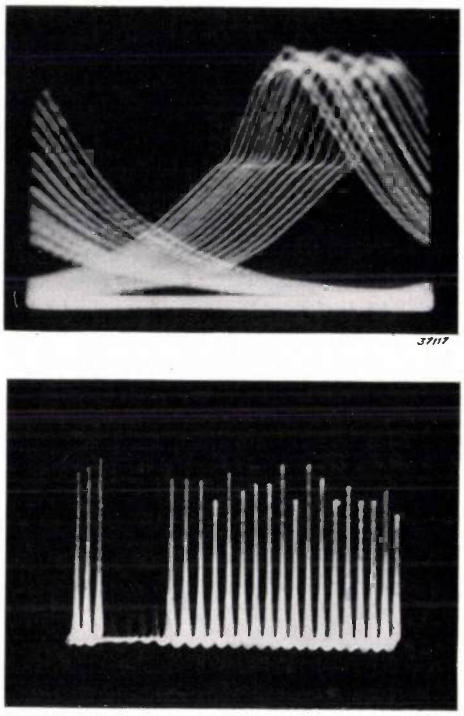
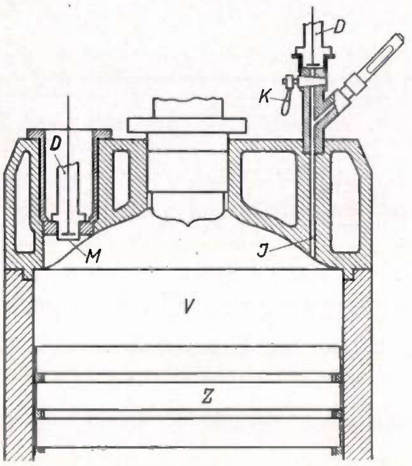
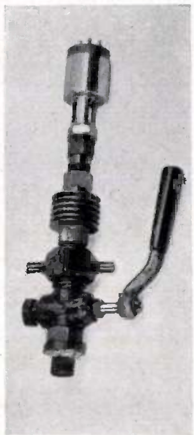


Fig. 5. *a*) Photograph of a series of successive pressure-time diagrams recorded without synchronization. By the gradual shifting of corresponding points of the diagrams the variation of the peak pressure and of the compression pressure is shown as a function of the time. *b*) A series of successive, very much compressed pressure-time diagrams. The variation of the top pressure can here also clearly be seen. The photograph of fig. 5*a* as well as those of fig. 6, 17 and 18 were kindly put at our disposal by the Testing Department of the *N.V. Werkspoor* of Amsterdam.

photograph of such a non-stationary image the variation with time of the peak pressure is clearly distinguishable, and also that of the compression pressure and the ignition, so that the engine constructor can judge on this basis whether there are irregularities in the combustion. Another possibility of obtaining a picture of the variation of the peak pressures is to make the time base very long, *i.e.* to make the fluorescent spot move only very slowly in a horizontal direction, so that a large number of (very compressed) diagrams are traced side by side, see fig. 5*b*.





a b

Fig. 6. If the membrane *M* of the pressure recorder *D* cannot be inserted into the cylinder wall itself (*a*, left) it must be connected with an indication channel *I* (*a*, right). In this channel gas vibrations as in an organ pipe may occur. *Z* piston, *V* combustion chamber, *K* indication tap. It must be mentioned incidentally that in this method of connection too great a heating of the membrane may sometimes also occur, since it cannot profit by the cooling by the (water-cooled) cylinder wall. In this case cooling fins must be constructed on the connection piece as may be seen in the photograph (*b*).

Vibrations on the expansion line of the diagram may also occur due to other causes than detonation. The pressure recorder often cannot be inserted directly into the wall of the cylinder, but must be mounted with a connecting piece on the indication channel (*fig. 6*) with which the cylinders of large engines are provided as a rule. Due to the sudden increase of pressure during the combustion (4-5 in fig. 2), characteristic vibrations can be excited in the gas column in the indication channel in the same way as in an organ pipe, which vibrations also appear in the pressure indicator diagram. By determining the frequency of the vibrations it is often possible to discover whether one is concerned with genuine detonation vibrations or with organ pipe vibrations. In the diagram of *fig. 7* for instance, it

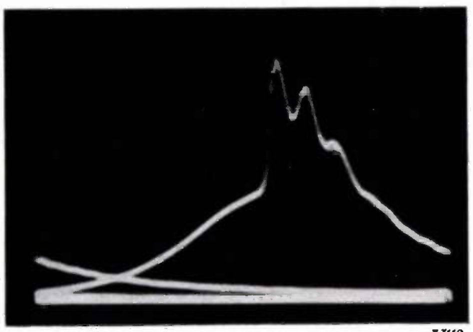


Fig. 7. Pressure-time diagram with vibrations on the expansion line, which, upon determination of the frequency, were found to be caused by organ pipe vibrations in the indication channel.

is found by measurement that the fundamental period of the vibration amounted to 0.037 times the duration of one revolution. The engine made 1600 r.p.m., the fundamental frequency of the vibration was therefore 720 c/s. On the other hand the length of the indication channel was 20 cm, and since in ordinary organ pipe vibrations the wave length of the fundamental tone is approximately four times the length of the pipe, in this case with a speed of propagation of the pressure waves in the combustible gases of 580 m/s a fundamental frequency of 580/0.8 = 725 c/s could be expected. The good agreement indicates that in this case it was probably a question of organ pipe vibrations.

Pressure-volume diagrams

The mechanical engineer will generally be more accustomed to record cylinder pressures as a function of the piston displacement than as a function of the time. Since the displacement of the piston from its highest position is proportional to the volume increase v of the combustion chamber, and $\int p\, dv$ is the work done by the gas or the recorded mechanical work, from the pressure-piston stroke diagram (pressure-volume diagram) the power delivered by the engine cylinder can be determined by planimetry. This was indeed originally the most important application of the indicator diagram.

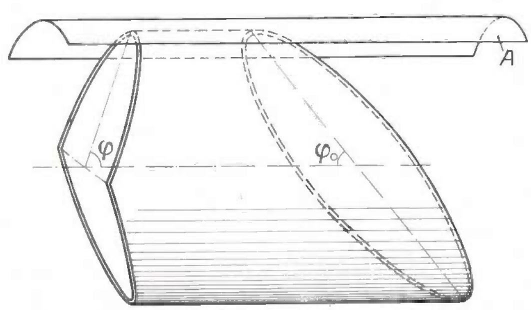


Fig. 8. Rotating cylinder condenser in the piston-stroke recorder. *A* fixed counter electrode.

In the article referred to [1]) the principle was briefly described of the arrangement whereby a horizontal deviation can be given to the fluorescent spot which is proportional to the displacement of the piston. The arrangement consists mainly of a cylinder cut off at the ends in a certain way, see *fig. 8*, which, together with a fixed counter electrode, forms a condenser. The cylinder is coupled with the crank-shaft, so that when the engine turns the capacity of the cylinder condenser varies periodically. The capacity variations are converted into voltage variations just as in the case of the pressure recorder, and the voltage variations are then fed to the horizontal deflection plates of the cathode ray tube. We shall here go somewhat more deeply into the construction of the cylinder condenser.

Since the counter electrode is only relatively narrow, it may be said that the capacity of the cylinder condenser at every moment is proportional to the length of cylinder at the point which is exactly opposite the middle of the counter electrode. Since the capacity variation must be proportional to the piston displacement, the mode of variation in the length of the cylinder as a function of the angle is hereby prescribed. With the help of *fig. 9*

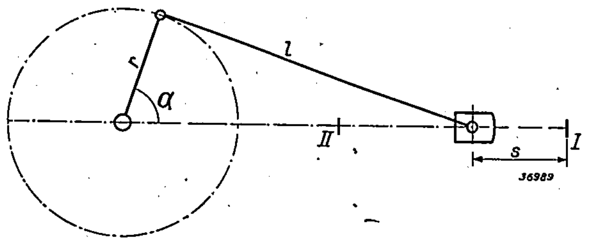


Fig. 9. Diagram showing the motion of the piston; r crank, l piston rod, I and II dead points. With infinitely long piston rod the piston displacement (s) would be sinusoidal as a function of the angle α of the crank. Due to the finite value of l a second harmonic enters the motion, whose amplitude depends upon the ratio $\varepsilon = r/l$.

for the displacement s of the piston from the highest position (dead point I, at which the volume of the cylinder is practically zero) the following formula is found:

$$s = r + l - r \cos \alpha - l\sqrt{1 - \varepsilon^2 \sin^2 \alpha} \quad . \quad . \quad (1)$$

The ratio ε between the length of the crank (r) and the piston rod (l) generally lies between 1/5 and 1/3.5. When the last term of (1) is developed in a series:

$$l\sqrt{1 - \varepsilon^2 \sin^2 \alpha} = l\,(1 - {}^1/_2\, \varepsilon^2 \sin^2 \alpha - {}^1/_8\, \varepsilon^4 \sin^4 \alpha - \ldots),$$

then the term with $\sin^4\alpha$ is already at least 100 times as small as the preceding term, so that we may write in sufficient approximation:

$$s = r - r \cos \alpha + (l\, \varepsilon^2 \sin^2 \alpha)/2$$

or. $$\frac{s}{r} = (1 - \cos \alpha) + \frac{\varepsilon}{4}\,(1 - \cos 2\alpha) \ldots (2)$$

The piston thus executes a practically sinusoidal motion upon which a weak second harmonic is superposed whose amplitude continues to depend upon the ratio of crank to piston rod of the engine in question. In *fig. 10a* the development of the cylinder condenser corresponding to equation (2) is drawn for the case where $\varepsilon = 0.222$ ($l = 4.5\ r$). The cylinder is bounded at one end according to the curve $1 - \cos\alpha$ and at the other by $(1 - \cos 2\alpha)\ \varepsilon/4$. The length of the intermediate section of straight cylinder is in principle a matter of indifference, since it involves a constant capacity upon rotation, while we are only concerned with the variation in capacity.

The first-mentioned boundary ($1 - \cos \alpha$) can be realized in a very simple way by cutting off the cylinder by a flat surface having an arbitrary slope φ_0 with respect to the cylinder axis. The boundary at the other end, according to $(1 - \cos 2\alpha)\ \varepsilon/4$ is, however, more difficult to construct practically. For the sake of simplicity in manufacture therefore the approximation given in fig. 10*b* is introduced. The curve $(1 - \cos 2\alpha)\ \varepsilon/4$ is replaced by $(1 - |\cos \alpha|)\ \varepsilon/2$, *i.e.* the cylinder is cut off by two flat planes whose position may be seen in fig. 8, while each plane makes an angle of φ with the cylinder axis, with $\cot \varphi = (\varepsilon/4) \cdot \cot \varphi_0$. It may easily be calculated that the greatest deviation between the curves of fig. 10*a* and *b* is equal to $\varepsilon/8$. For $\varepsilon = 0.222$ this

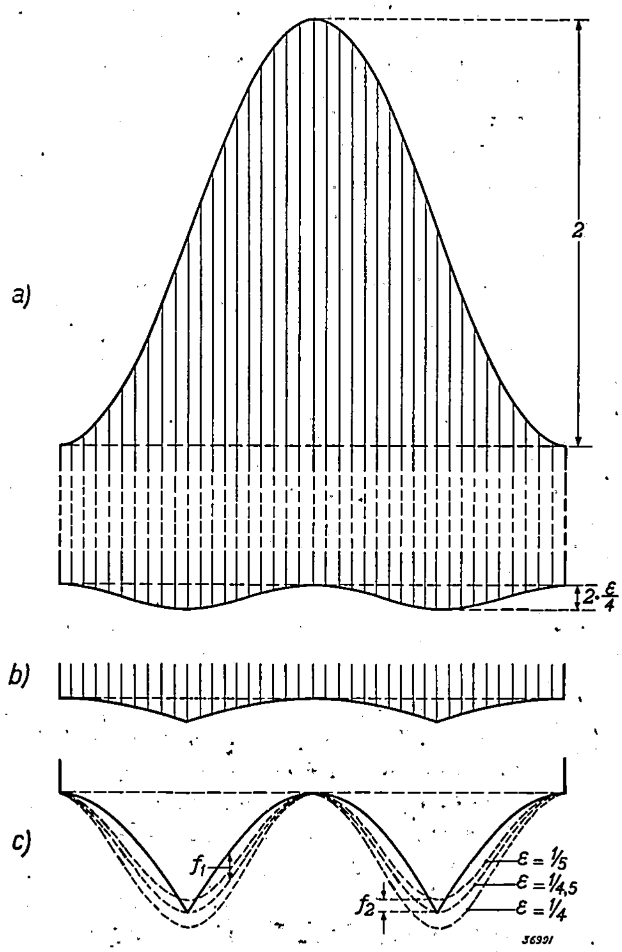


Fig. 10. Development of the condenser cylinder: *a*) theoretically desired form for a crank-piston rod ratio $\varepsilon = 1/4.5$; *b*) approximation of the desired form, chosen because of its easy realization; *c*) on a scale 5 times as large: shape used (full line, like *b*) and theoretically desired shapes (broken lines) for different values of ε.

is a shift of the abscissa of 1.4 per cent of the total amplitude, an error which may be permitted without serious consequences.

If the same piston-stroke recorder is used for engines with a different ratio ε of crank to piston rod, larger errors may occur. This may be seen directly in fig. 10*c* where in addition to the curve for the recorder used (as in 10*b*) the form of the desired curve is drawn for several values of ε. In *fig. 11* the maximum positive or negative deviation is plotted as a function of ε. If an error of 2.5 per cent is allowed, the piston-stroke recorder which is constructed with $\cot\varphi = 1/4 \cdot 0.222 \cot\varphi_0$ is found to be still usable for engines with $1/8 < \varepsilon < 1/3.6$. Values of ε outside this range pratically do not occur.

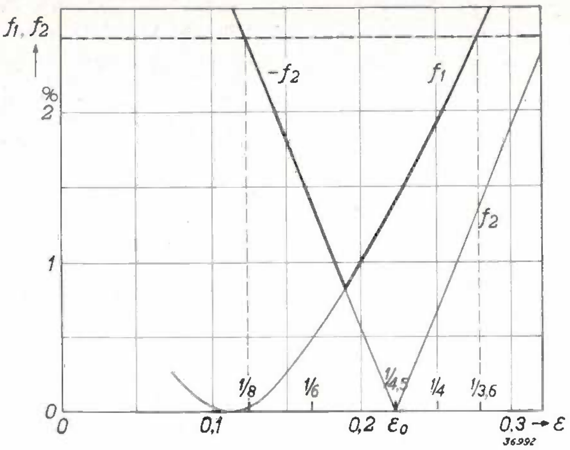


Fig. 11. Maximum difference f_1, f_2 between the deflection obtained and that desired of the cathode ray, in per cent of the total piston stroke, as a function of the crank-piston rod ratio ε of the engine being tested. The values of $\cot\varphi$ and $\cot\varphi_0$ (see fig. 8) are here chosen in the ratio of 1 : 18. If a maximum positive or negative error of 2.5 per cent is allowed, the piston-stroke recorder so constructed can be used for all engines with $1/8 < \varepsilon < 1/3.6$, as the figure shows.

The above considerations actually hold only for the case where the counter electrode is infinitesimally narrow. When it has a finite width (angle Θ), then the capacity at every moment is given by the average length of the cylinder in the effective sector of the surface of the cylinder. The effect of this is [3]) that higher harmonics in the variation of the cylinder length are weakened, the n^{th} harmonic by a factor $(\sin\Theta/2)\,(n\Theta/2)$. In order to obtain sufficient capacity Θ had practically to be made equal to 10°. The second harmonic hereby experiences only a relative weakening by a factor 0.995, so that the effect may be neglected.

In *fig. 12* a normal pressure-volume diagram is reproduced, recorded with the help of the rotating cylinder condenser described. In order to obtain such a diagram the counter electrode of the cylinder condenser must have a position such that the highest position of the piston corresponds to the smallest

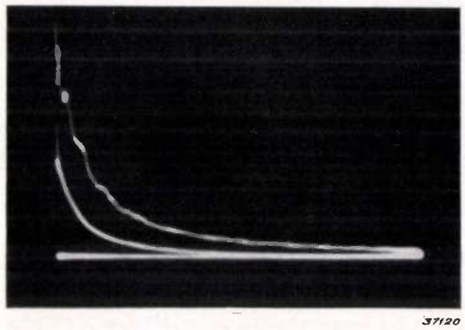


Fig. 12. Normal pressure-volume diagram recorded with the pressure indicator GM 3 154 and the piston-stroke recorder described. (Here again gas vibration in the indications channel are superposed.)

capacity (deflection of the cathode ray zero, or equal to a given initial deflection). From the diagram by planimetry, as already stated, the power delivered by the engine cylinder can be determined, and from this the mechanical efficiency can be calculated, for example with the help of the power measured at the crank-shaft.

Since in the neighbourhood of its highest position (and lowest position) the piston moves relatively slowly, the important processes of ignition and combustion, which take place about the moment when the piston is in the highest position, are compressed in the *p-v* diagram into a short section of the abscissa. Peculiarities and possible deviations in the combustion cannot therefore be easily distinguished in the normal *p-v* diagram. At the time when mechanical indicators were generally used the following device was employed to make up for this unpleasant lack. The motion in the direction of the abscissa was shifted 90° in phase with respect to the actual piston movement, so that the successive values of the pressure in the combustion chamber were not recorded above the corresponding volume values in the diagram, but were

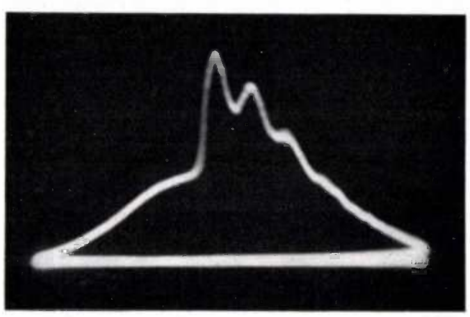


Fig. 13. "Shifted" pressure-volume diagram (same as fig. 12). The counter electrode of the cylinder condenser fig.8 was here rotated 90° with respect to its normal position.

3) The effect is quite similar to that of the finite width of a scanning slit which was discussed in: J. F. Schouten, Synthetic sound, Philips techn. Rev. **4**, 167, 1939 (see especially page 169).

shifted a quarter period. The combustion pressures thus lay in the middle part of the abscissa, where the motion is most rapid. The diagram obtained in this way, the so-called shifted *p-v* diagram, of which an example is given in *fig. 13*, gives a better idea of the actual combustion process than the normal *p-v* diagram, and could be obtained with the mechanical pressure indicator simply by moving a lever.

For the mechanical engineer who is accustomed to work with these shifted *p-v* diagrams, it was very simple in the case of the electrical indicator to obtain such diagrams. It was only necessary to rotate the counter electrode of the above-described piston-stroke recorder through the desired angle. The possibility of such a rotation had in any case to be provided for in connection with the testing of different cylinders of the same engine whose cranks always stand at different angles. The construction of the piston-stroke recorder shown in *fig. 14* is such that the counter electrode can be turned with the hand and set at intervals of 30°. The most commonly occurring crank angles are multiples of 30°. At the same time a contact is also introduced on the axis of the rotating condenser which is closed once per revolution and which, as described above, serves for the synchronisation in the recording of pressure-time diagrams.

For routine testing of large engines, for instance of large Diesel installations, it is important to be able to test each cylinder separately. The indicating instrument is therefore provided with several connections which, *via* several cables, are connected to pressure recorders on the different cylinders. When pressure time diagrams are being recorded, the diagrams of the different cylinders can be made to appear successively on the screen of the cathode ray tube simply by operating a switch. When, however, the piston-stroke base is being used, in addition to switching over to the corresponding pressure recorder, the counter electrode of the piston-stroke recorder must also be brought into the correct position for each cylinder.

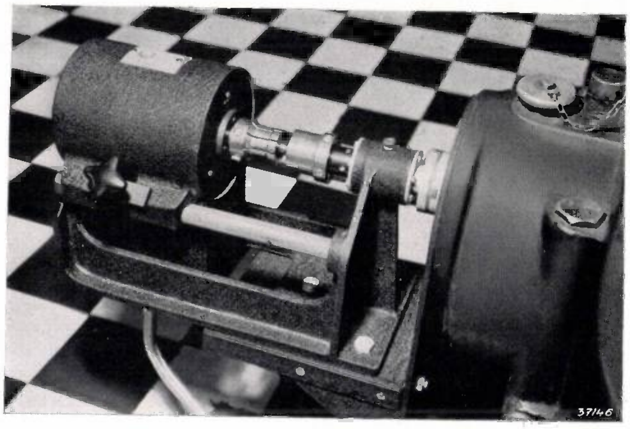


Fig. 15. Piston-stroke recorder (GM 4 300), larger model with operation at a distance, for large Diesel installations. The rotating condenser can be coupled with the crank-shaft by means of a sliding coupling arrangement which is here opened. It is so made, that it can be closed when the shaft is turning and the condenser cylinder automatically assumes the correct position with respect to the crank angles.

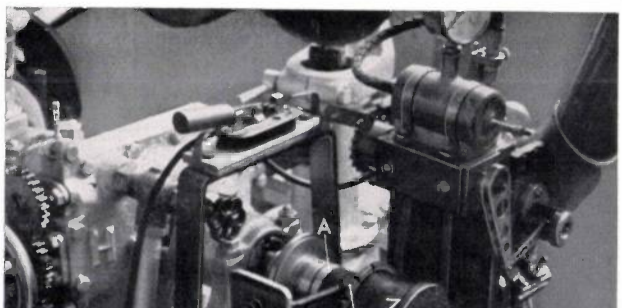


Fig. 14. Piston-stroke recorder *Z* (GM 4 301) coupled with the crank-shaft of a high-speed engine. *T* arrangement for fixing the counter electrode at definite angles, *S* connection for the synchronisation, *A* coupling with the engine shaft, *K* cable connection.

In order to simplify this manipulation a special piston-stroke recorder has been developed in which the counter electrode can be operated at a distance with the help of a small servo motor. *Fig. 15* is a photograph of this recorder, while *fig. 16* shows the construction of the indicator which is used in combination with it for large Diesel installations. The counter electrode is driven by the servo motor *via* a kind of Maltese cross which makes the electrode stop for a moment at intervals of 30°, so that the adjustment on the cylinders with different crank positions becomes much easier. A contact disc is attached to the counter electrode, which causes a series of signal lamps on the indicating instrument to light up, so that the position of the counter electrode can continually be checked. At the same time in this model of the piston-stroke recorder the axis of the rotating condenser has a contact disc by means of which the cathode ray can be periodically suppressed in such a way that the diagram on the fluorescent screen exhibits an interruption of the line every 20°. This provides easier orientation in the diagram.

In connection with the satisfactory functioning of the latter contact disc, the crank-shaft may not make more than 800 r.p.m. With large engines, however, for which this model is chiefly intended, such a high speed of revolution practically never occurs. The simpler model shown in fig. 14 of the piston-stroke recorder can be applied up to much greater speeds, and therefore to high-speed engines such as racing engines.

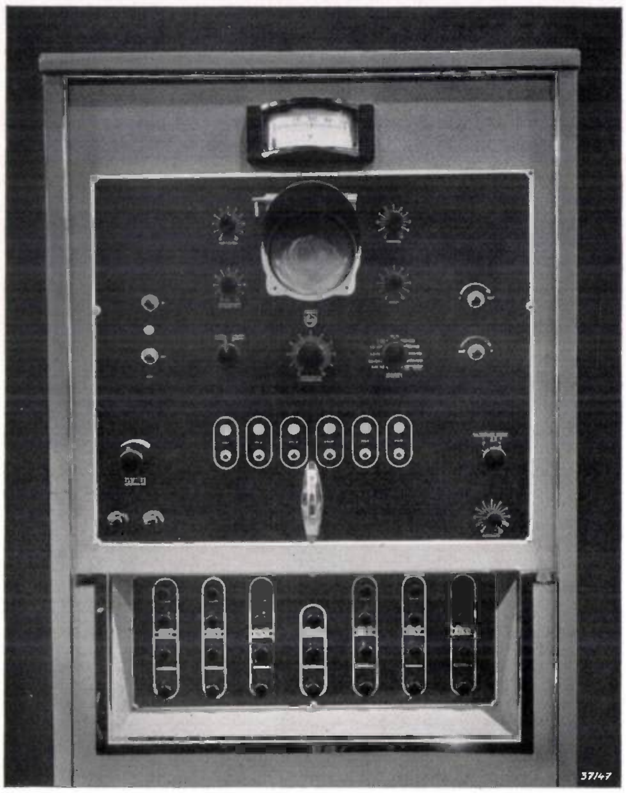


Fig. 16. Pressure indicator, model for large installations. In the centre of the upper half is the screen of the cathode ray tube. The middle of the seven sets of knobs visible below serves for the regulation of the compensation and the amplitude for the piston-stroke base (see the article referred to in footnote [1]); the other six sets to the right and left serve for the corresponding regulations for six pressure recorders on different cylinders of the engine. Above these knobs may be seen a row of six times two signal lamps which light up when the counter electrode of the piston-stroke recorder for the corresponding cylinder is in the normal position or rotated 90°.

The needle-stroke diagram

In addition to the diagrams discussed, the so-called needle-stroke diagram which records the motion to the fuel injection needle is also of importance to the constructor of Diesel engines. This needle is opened by the fuel pump operated by cams on a shaft coupled with the crank-shaft. The position and shape of the cams must be so chosen that the opening and closing of the valve needle takes place at the correct moment; furthermore the fuel supply line must be of the proper size so that the periodic pressure increase in the fuel oil will be propagated in the desired way from the pump to the needle.

In the article repeatedly referred to [1]) it was explained that the pressure recorder there described can in a simple way be adapted to the recording

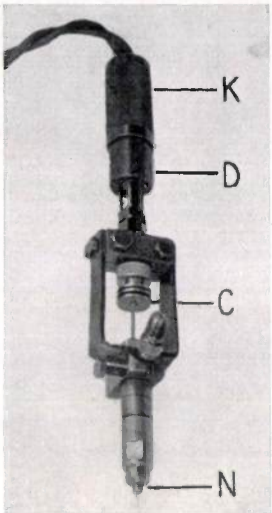


Fig. 17. Arrangement for the recording of a needle-stroke diagram. N fuel needle, C variable condenser of the pressure recorder D used as vibration recorder, K connecting piece for the cable.

of mechanical vibrations. The arrangement amounts to the setting up opposite the object to be investigated of a fixed, electrically insulated counter electrode, whose capacity with respect to earth varies due to the vibration. These capacity variations, in the same way as in the piston-stroke recorder and the normal pressure recorder, are converted into voltage variations which can be used for the deflection of the fluorescent spot of a cathode ray tube.

For recording the needle-stroke diagram a plate is now fastened above the needle, and above that the counter electrode of the vibration recorder is placed, see *fig. 17*. As horizontal deviation for the

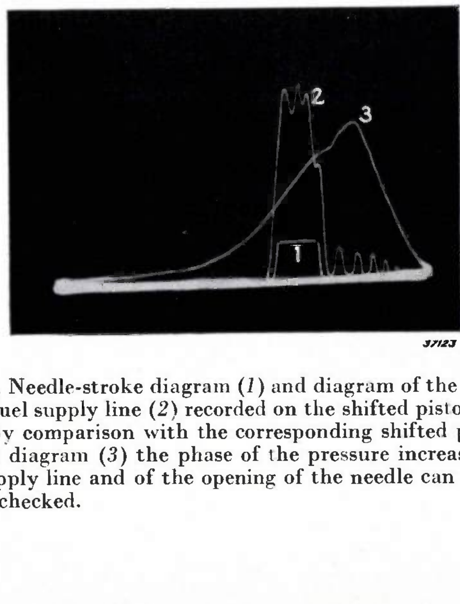


Fig. 18. Needle-stroke diagram (*1*) and diagram of the pressure in the fuel supply line (*2*) recorded on the shifted piston-stroke base. By comparison with the corresponding shifted pressure-volume diagram (*3*) the phase of the pressure increase in the fuel supply line and of the opening of the needle can be accurately checked.

needle-stroke diagram the "shifted piston-stroke basis" is generally used, which was described above. If the shifted pressure-volume diagram itself is recorded simultaneously, it is easy to find out whether the injection takes place at the correct phase of the piston movement. The use of the shifted piston-stroke basis is here called for, since the injection takes place just in the neighbourhood of the dead point of the piston, where the normal piston-stroke basis most strongly compresses the diagram.

In *fig. 18* such a needle-stroke diagram is shown, together with the corresponding shifted pressure-volume diagram. The phase of the moment of injection can be accurately determined to within several crank degrees. In this figure, on the same basis, a diagram is also given which was recorded of the pressure variation in the fuel supply line at the pump. This liquid pressure can be measured with the ordinary pressure recorder used for gas pressures or with specially developed pressure recorders (with larger or smaller measuring range). In fig. 18 therefore use is made of all three possibilities of application of the indicator GM 3 154, namely for the recording of gas pressures, liquid pressures and mechanical vibrations.

TEMPERATURE MEASUREMENTS WITH THE OPTICAL PYROMETER IN THE HARDENING DEPARTMENT

by J. RIEMENS.

536.52

A pyrometric method is described of measuring the temperature of a liquid bath between 850 and 1 450 °C with an accuracy of $\pm 2°$. This arrangement is used in the factory for the measurement of the temperature of salt baths in the hardening department.

The hardening of steel is for the purpose of giving the metal that structure which possesses the desired properties. In earlier years when this treatment was carried out one could for the most part rely upon the experience of the foreman of the department. Modern metallurgy, however, has fundamentally altered the situation. New kinds of steel for special purposes have been developed, and these steels require a very precisely determined heat treatment in order to attain the desired properties. This is especially true of high speed tool steel which is much used for tools for metal working. In order to harden these tools they are introduced into a bath of fused salts and then cooled in a special way.

The temperature of the salt bath is of the greatest importance. In order to obtain reproducible results this temperature may not deviate more than about 5 degrees from the value at which the optimum result is obtained. This value generally lies between 850 and 1 450 °C.

Temperature measurements in the region mentioned can be carried out with sufficient accuracy not only with the help of thermoelements but also with the help of an optical pyrometer. For use in this temperature range only those ordinary thermoelements can be used which consist of platinum and platinum rhodium. In the practical use of these elements it has been found that the thermoelectric force gradually depreciates upon repeated heating above 1300 °C, so that one is compelled to calibrate the element repeatedly and to replace it by a new element when the thermo electric force has diminished too much. This objection led to the adaptation of the optical method of measuring.

Optical pyrometry is based upon the fact that the brightness of a hot surface, for instance that of the salt bath, increases rapidly with the temperature. The brightness of the salt surface at a given temperature, however, also depends upon the composition of the salt, and small impurities may play an important part, so that the required measuring accuracy is difficult to obtain.

Greater accuracy is obtained by immersing in the salt a body whose brightness is accurately known as a function of the temperature. An absolutely black body satisfies the requirement best. The black body can be realized by immersing in the bath a tube closed at the lower end. The part immersed must have a length at least four times the diameter of the tube. The radiation which emerges vertically out of the opening of the tube is then exactly the same as that of a black body of the temperature of the salt bath, within several tenths of a per cent. The temperature of the salt bath can therefore be determined directly by measuring the brightness of the opening of the tube.

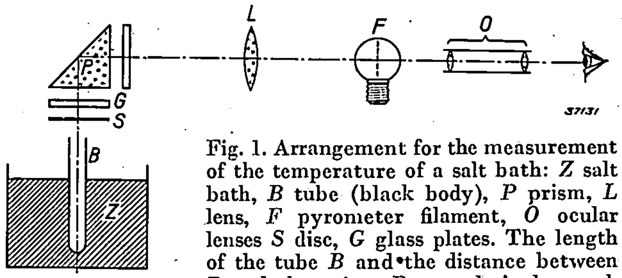


Fig. 1. Arrangement for the measurement of the temperature of a salt bath: Z salt bath, B tube (black body), P prism, L lens, F pyrometer filament, O ocular lenses S disc, G glass plates. The length of the tube B and the distance between B and the prism P are relatively much greater than shown in the figure.

In *fig. 1* the arrangement is given. The bottom of the tube is focussed by means of the prism P and the lens L on a plane in which a glowing wire, the pyrometer filament, is situated. With the help of the ocular lenses O the pyrometer filament and the image of the bottom of the tube are examined together. If the current through the pyrometer filament is so regulated that the filament exhibits the same brightness as the tube bottom (and thus becomes invisible), the deviation of the ammeter which indicated the pyrometer current immediately furnishes a measure of the required temperature.

In order to obtain a correct result it is found that extreme care must be taken that salt vapours do not interfere with the pyrometry. By closing the tube this is sufficiently well ensured. Nevertheless in the long run the danger remains that small amounts of salt will be deposited on the optical parts of the pyrometer and thus change the calibration. In order to prevent this a metal disc S is introduced between the tube and the prism, and is only opened just before the measurement. For all security the prism is also protected by two removable glass windows G which can be wiped clean from time to time. The prism therefore need never be cleaned, and cannot therefore be brought out of adjustment.

After some practice in the setting of the optical pyrometer an accuracy of measurement of about 2 °C is attained at a temperature of from 1 300 to 1 450 °C. An instrument of very good quality must, however, be used for measuring the pyrometer current. If an tungsten wire of 75 microns diameter is used as pyrometer filament, for example, it is found that the current must be able to be measured with a reproducibility of 2 tenths per cent.

Since the setting up of a precision instrument which satisfies such heavy requirements meets with difficulty in the rough surroundings of the hardening department, a kind of compensation connection

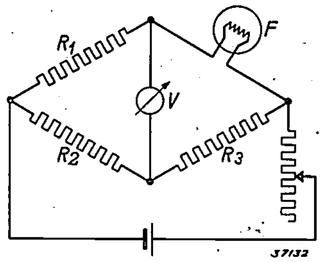


Fig. 2. Compensation connections for an accurate determination of the pyrometer current. R_1, R_2, R_0 bridge resistances, F pyrometer filament.

is used for the measurement of the pyrometer current. The filament of the pyrometer is connected in one of the branches of a resistance bridge (see *fig. 2*). The resistance of the filament changes approximately proportionally with its temperature. The resistances of the other branches of the bridge are left unaltered; they are so chosen that at a temperature of 850 °C the bridge is balanced, while at a temperature of 1 450 °C the full deviation of the meter V is obtained.

These connections have the advantage that the temperature range of practical importance occupies the whole scale of the meter, while with a direct measurement of the pyrometer current only about half of the scale of the ammeter can be used for this temperature range. The result is that the same accuracy of ± 2 °C can be reached with a reproducibility which is 2.5 times as poor. With a good switchboard instrument one can measure accurately to within 5°, which is enough for practical purposes.

The constancy of the optical pyrometer is extremely great. Thanks to the relatively low filament temperature of 1 540 °C at the highest, the rate of evaporation of the tungsten wire is still very low, so that the properties of the wire during use remain practically unchanged. A semi-annual check of the instrument therefore gives sufficient guarantee that the permissible tolerance in the measuring result is not being exceeded.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS GLOEILAMPENFABRIEKEN

1506: A. A. Kruithof: Townsend's ionization coefficients for neon, argon, krypton and xenon (Physica **7**, 519-540, June 1940).

Townsend's ionization coefficient was determined for krypton and xenon as a function of E/p_0. Together with results already published for neon and argon (see **1106** and **1206**), the measurements show that for $E/p_0 < 40$ V/cm $\times$ mm the ionization is strongest in neon. It is much less strong in argon and krypton and weakest in zenon. For high values of E/p_0 the order is exactly reversed. Beginning with the fact that the average length of path as a function of the electron energy varies in about the same way for krypton and xenon as for argon, the ionization and probability of excitation could be approximately calculated for krypton and xenon from those of argon, with the help of the results of the measurements.

Furthermore the number of electrons was determined as a function of E/p_0 which on an average, per positive ion formed in the gas, are freed from a copper cathode. The curves found make it possible to divide the freed electrons into two groups. One group contains the electrons liberated from the cathode by the collision of positive ions, the other group contains the photoelectrons freed by the very short-wave ultra violet radiation of the gas. For high values of E/p_0 the first group is the largest, for low values the second.

1507: J. D. Fast: The action of gases on solid metals (Chem. Wbl. **37**, 342-350, June 1940). (Original in the Dutch language)

In the corrosion of solid metals by gases a solid solution of the gas in the metal can be formed, whereby the gas atoms diffuse toward the interior of the metal, and a solid reaction product may be formed on the surface whereby the metal atoms diffuse through the layer formed in the direction of the gas phase. These phenomena are discussed in detail and illustrated with numerous examples, some from the literature and some from the author's own experiments. Part of the material discussed in this article will be found in the August number of this periodical (**5**, 217, 1940: Metals as getters).

1508*: F. A. Kröger: Luminescence in solids containing manganese (Dissertation, Amsterdam, July 1940).

In order to gain insight into the part played by manganese as an activator in phosphors, silicate and sulphide phosphors were investigated. In both, the manganese is present in solid solution in the crystal lattice, and is equivalent to the metal ions of the substance which serves as basic material. The light which excites the phosphors corresponds to their absorption spectrum and consists of discrete bands in the ultra violet and visible spectrum (characteristic of manganese), of a broad sharply defined band in the ultra violet resulting from the fundamental absorption of the built-in manganese compound, and in the third place of a broad band, at the long wave end bounded by the above-mentioned and at the short wave end with no observable boundary. This latter band is the fundamental absorption band of the basic material. Irradiation with light in the two last mentioned parts of the spectrum leads to fluorescence and phosphorescence accompanied by photoconductivity. Irradiation in the first part gives only fluorescence. The luminescence is characteristic of the built-in manganese ions and is ascribed to electron transitions within the manganese ion (transitions between the fundamental terms of the d^5 configuration of the half-filled d shell).

The zone theory, which takes account of the possible energy levels in solid substances, is extended to include solid solutions, so that a usable model for the phosphors is obtained. With this it is shown that according to the characteristics of the emission a classification of the phosphors into two groups is possible. In the first group the emission process takes places within the excited centre, the emission is only secondarily dependent upon the environment. In the second group this environment, the basic material into which the activator is built, plays a part and the emission spectrum is strictly determined thereby. Manganese phosphors, according to the investigations referred to above, belong to the first group.

*) An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

VOL. 6 No. 2, pages 33-64 FEBRUARY 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

AERODROME ILLUMINATION BY MEANS OF WATER-COOLED MERCURY LAMPS

by Th. J. J. A. MANDERS. 628.971.8 : 656.71 : 621.327.3

For the illumination of the landing area of an aerodrome a light source set up at the edge of the area is used which emits a flat, fan-shaped beam of light. Rough calculation shows that the beam must have a vertical spread of the order of 1° or less, while the luminous intensity must amount to more than 1 million candle power. For the realization of such a beam, water-cooled high-pressure mercury lamps are particularly suitable. A description is given of the mercury lamp SP 2 000 W, which for the purpose in view is provided with a cooling system with a closed water circulation.

For safety in making night landings on an aerodrome, in addition to an adequate beaconing of points around the aerodrome [1]), the first requirement is an efficient illumination of the landing area itself. From the nature of the case this area must, however, be quite free of obstacles such as masts and the like. The illumination, which naturally involves the setting up of apparatus at a certain height above the ground, can therefore only be installed along the edge of the landing field. The requirements which result from this condition for the construction of the light sources to be used will be briefly discussed in this article. It will be found that the high-pressure mercury lamp, particularly that with water cooling, is especially suitable for this purpose. Such a lamp provided with a closed cooling system will be described in detail.

Requirements for the illumination of the landing field

It is in general desirable that such a length of the landing area should be illuminated that the aeroplane upon landing and running out comes to a stop upon the lighted part of the field, and that such a width of the field should be illuminated that the pilot has plenty of leeway in choosing the spot where he shall land, and that there shall be no disturbing contrast to the right or left of his field of vision. According to the internationally established rules (C.I.E. 1935 and 1939) on the basis of these considerations minimum dimensions of 600 × 300 m are recommended for the part to be illuminated.

Since the pilot must not be blinded by the source of light, the illumination is arranged to be from only one side of the field, namely in such a way that the direction of the light rays is almost coincident with the direction in which the aeroplane moves in landing. Furthermore, in order not to form an obstacle and also for practical reasons, the light source may not be too high. An arrangement is thus arrived at like that sketched in *fig. 1*. The light

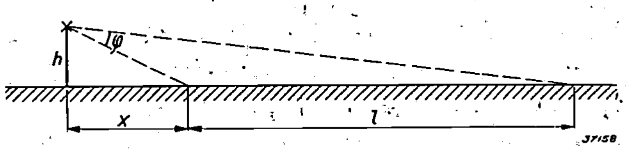


Fig. 1. Diagram of the arrangement of a landing light. φ vertical angle of spread of the beam. The height h is here very much exaggerated for the sake of clearness. Drawn to scale, h would be about $^1/_4$ mm in this figure.

source here throws a flat, fan-shaped beam on the field [2]).

In the above-mentioned international rules it is required that the lowest illumination intensity occurring, *i.e.* that at the farthest extremity of the part to be illuminated, must amount to 1.5 lux

[1]) See G. L. van Heel, The illumination and beaconing of aerodromes, Philips techn. Rev. 4, 93, 1939.

[2]) The oblique illumination of the field causes long shadows to be cast by irregularities of the surface. In order to limit this undesired effect the light source will be set up in every case as high as is compatible with the objections mentioned.

(measured in a vertical [3]) plane). It is still better to count on a minimum of 2.5 lux. On the other hand it is recommended that the highest illumination intensity occurring, which is found at the side of the field near the light source, should be limited to 25 lux, in order not to spoil the adaptation of the pilot's eyes. In order to satisfy this requirement the light source is mounted at a sufficient distance (x in fig. 1) from the edge of the area to be illuminated.

If for the sake of simplicity we assume that the light beam is sharply bounded above and below and that the luminous intensity is constant over the whole vertical spread, we can easily calculate the required light intensity I and the angles of spread of the beam. When l is the length of the area to be illuminated ($l = 600$ m) the distance x follows from the condition that

$$\frac{I}{x^2} : \frac{I}{(l+x)^2} = 25 : 2{,}5,$$

$$x = 0{,}46\ l \approx 280 \text{ m}.$$

The light intensity thus becomes

$$I = 25 \cdot 280^2 \approx 2 \text{ million candle power.}$$

Between the height h of the light source above the ground and the vertical angle of spread φ the following relation is found:

$$\varphi \approx \frac{h}{x} - \frac{h}{l+x} \approx 0{,}0025\ h.$$

At a height of $h = 3.5$ m, $\varphi \approx {}^1/_2{}^0$. If the usual condition is made that the pilot must also be able to land at angles of up to 45° to the right or left of the direction of the beam, it follows (see *fig. 2*) that the horizontal spread of the beam must be at least 80°.

Actually of course it is impossible to create a beam of light with the sharp boundaries here assumed. The light intensity will decrease more or less gradually toward the edges. We have also entirely neglected atmospheric absorption in the above considerations. Nevertheless, it is sufficiently clear what unusual requirements are here made of the light source.

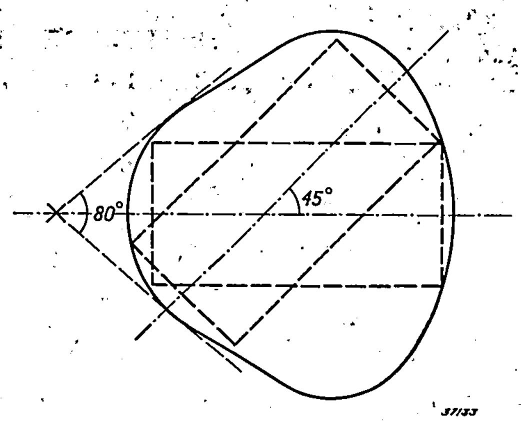


Fig. 2. Ground plan of the landing area upon which the illuminated part is indicated. Approaching in any arbitrary direction within an angle of 90° the pilot must have an illuminated area 300 × 600 m.

The most suitable method of obtaining the fan-shaped beam described is to place a lamp along the horizontally placed axis of a cylindrical parabolic mirror, see *fig. 3*. In the direction of the axis the luminous body may be fairly long. In order to keep the vertical spread of the beam small, however, the transverse dimensions of the radiating body must be small with respect to the dimensions of the mirror. If huge mirror constructions are not desired the lamps must be constructed with a very narrow luminous surface, while in order to obtain the necessary light intensity, the brightness of the radiating surface must be very high. It is possible to satisfy these requirements with suitably constructed electric filament lamps; *fig. 4* shows the filament systems which must be used. A more elegant solution of the problem, however, lies in the use of high-pressure mercury lamps which naturally have a linear form and a great brightness. The spectral composition of mercury light is also especially suitable for illuminating a green grass field with which most aerodromes are covered [4]) (either completely or surrounding the starting runways).

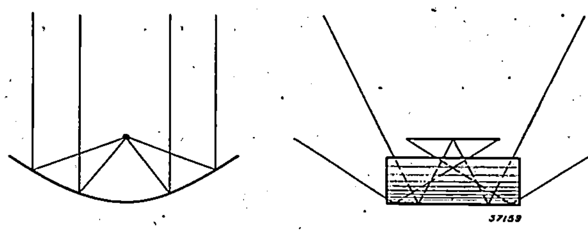


Fig. 3. Cylindrical parabolic mirror at whose axis the light-emitting body is situated.

[3]) In connection with the oblique illumination the measurement on a vertical plane is called for. These measurements, however, usually also provide a good measure for the effect obtained (the brightness observed by the pilot), since the landing areas are usually covered with grass, so that the illuminated "plane" actually consists of numberless, almost vertical reflecting planes. The greatest brightness is actually observed in the direction of the illumination. The level of brightness then corresponds approximately to that on well-lighted roads with an illumination intensity of 3 to 10 lux.

[4]) See in this connection the article referred to in footnote [1]).

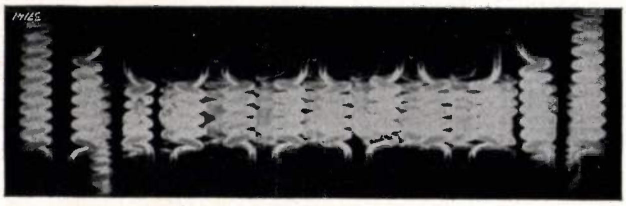


Fig. 4. In order to obtain the desired beam formation the light-emitting body must be made very narrow, while it may of course be long. The photograph shows the constructions used when the necessary light flux is to be obtained by means of a filament.

Air-cooled and water-cooled mercury lamps

In the article referred to [1]) a description has already been given of a high-pressure mercury lamp (HP 1000 W), which was developed for use in aerodrome illumination. With the help of a cylindrical parabolic reflector a maximum light intensity of 25 000 candle power was obtained, with an angle of spread of the beam of about 2° (by the angle of spread of the beam is meant the angle at which the light intensity has fallen to one half the value occurring in the direction of the axis). In order to obtain the desired illumination of the landing area, therefore, several of these lamps must be used.

High-pressure mercury lamps have a greater brillance, the higher the power consumed per cm length of the discharge column [5]). Not only does the light flux increase with this power, but in addition the efficiency also rises, while the diameter of the light-emitting discharge decreases. Since the power which the high-pressure mercury lamp can consume is mainly limited by the heating of the tube wall, considerable improvement in brightness is obtained by applying a forced cooling of the tube wall. While in the case of the lamps type HP 1000 W mentioned, where the cooling had to take place by means of the surrounding air, a maximum brightness (at the axis of the discharge) of 1400 c.p./cm² occurs, when water cooling is employed values 20 to 30 times higher are easily obtained, and in laboratory tests values even 100 times as high [6]). In *table I* the various data are given for the water-cooled mercurylamp SP 2000 W, which, like the HP 1000 W, is especially intended for aerodrome illumination. It may be seen that even with one or two lamps the whole landing area can be illuminated. In *fig. 5* the vertical and horizontal light distribution curves of the lamp with reflector are reproduced, while *fig. 6* shows the distribution of the illumination intensity over the illuminated area.

Table I

Data of the water-cooled mercury lamp SP 2 000 W

Power consumed	2 000 W
Light flux	ab. 120 000 lm
Efficiency	ab. 60 lm/W
Length of discharge	50 mm
Greatest brightness (at axis)	ab. 30 000 cp/cm²
Average life	150 hrs
Beam obtained (with reflector):	
vertical spread	ab. 1,1°
maximum light intensity	ab. $1{,}3 \cdot 10^6$ cp.

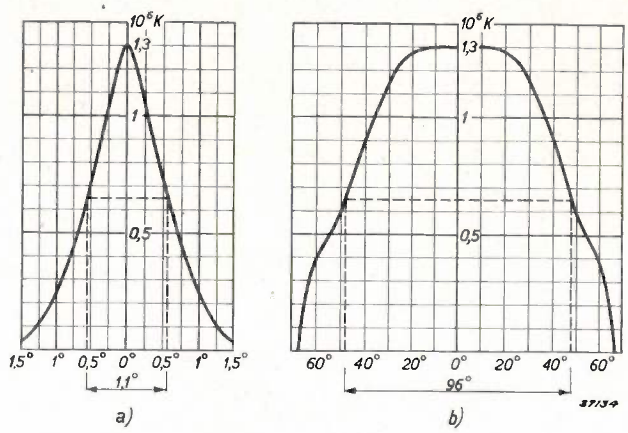


Fig. 5. Light distribution curve of the lamp SP 2 000 W in a cylindrical parabolic mirror; *a*) in a vertical plane perpendicular to the focal, line, *b*) in the horizontal plane through the focal line. The light intensity at the axis of the beam is 1.3 million candle power, the angle of spread in the vertical plane is about 1°, and in the horizontal plane about 90°.

Besides the economic advantages which result from the greater brightness (possibility of a better beam formation) and the higher efficiency, the water-cooled lamp has another important technical advantage over the lamp which is not artificially cooled. When a mercury lamp is switched on it must first become warm before it produces the full light flux. When the lamp is switched off, re-ignition is only possible after sufficient cooling of the lamp (decrease of the mercury pressure). While the lamp with natural cooling requires several minutes both for warming up as well as for cooling, the water-cooled lamp after being switched off is almost immediately ready for action again, and upon switching on, thanks to the high power consumed, it heats up to final temperature within a few seconds. These advantages are very important for an intermittent use such as will often be necessary on aerodromes.

[5]) See for example Philips techn. Rev. **2**, 165, 1937 and the articles there referred to.

[6]) In this way it has been possible to exceed the brightness of the sun which is about 165 000 c.p./cm² (Philips techn. Rev. **1**, 62, 1936). In the article referred to in footnote [1]) the lamp described in the following, SP 2 000 W, which was then still at an experimental stage, was announced.

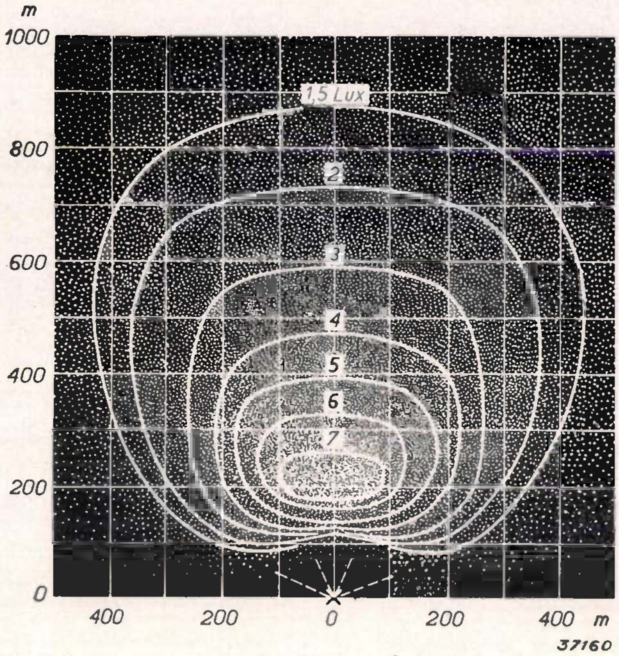


Fig. 6. Diagram of the distribution of the light over the illuminated landing area.

Construction of the water-cooling system

The way in which the water-cooling system is constructed is shown in *fig. 7* [7]). The quartz tube filled with mercury vapour is situated along the common axis of two cylindrical glass jackets. The water first flows through the inner cylinder and then flows back through the ring-shaped space between the two glass cylinders. The inlet and outlet for

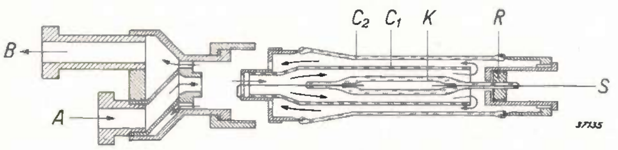


Fig. 7. Mercury lamp with water cooling. The water flows in at A, along the quartz tube K of the lamp inside the inner glass cylinder C_1, then between the cylinders C_1 and C_2 back again, and out at B. On the right the current lead S which is insulated from the cooling water by means of the rubber cap R which is pressed into place by a "Philite" cap.

the cooling water are therefore both at the same end of the lamp, and the connection is made by means of the connecting piece shown is fig. 7. This forms at the same time the current lead to one electrode, while the current lead of the other electrode is insulated from the cooling water by means of a rubber cap over the end of the quartz tube (see figure). The connections for the current at one end and the inlet for the water at the other are so constructed and mounted in the reflector that when the lamp with jackets is adjusted, the light-emitting discharge automatically coincides with the focal line of the parabolic mirror.

An effective cooling is obtained by the double stream of water. While the light must indeed pass through a double layer of glass and water here, only a small amount of the visible light is absorbed, so that it constitutes no objection in this case. The necessary amount of cooling water with a permissible rise in temperature of the water of 70° amounts to 6 to 8 l/min. Due to the relatively small surface on which the transfer of heat can take place, fairly high speeds of flow must be maintained (about 5 m/sec), for which an excess pressure of 0.3 to 0.5 atmosphere is required.

From what source shall the necessary cooling water be taken? It would seem obvious that it should be taken from the local mains. For various reasons, however, this apparently simple solution is not satisfactory. In the first place, the position where the light source must stand for a landing is not usually fixed. The aeroplanes must in general land against the wind; therefore according to the direction of the wind the landing light must stand at different points along the edge of the landing area. Whether this condition is met by setting up a large number of lights around the aerodrome, from which a choice can be made, or by means of one movable unit, — in either case an elaborate network of water pipes would have to be laid for the cooling water. A second objection is encountered in the difficulty of protecting the pipe lines from freezing in the winter. And last but not least, the water from the mains is generally not pure enough, so that a deposit is quickly formed on the quartz tube of the mercury lamp. This deposit consists partly of the familiar boiler scale, often, however, of carbon and other substances which are formed by the decomposition of organic and other impurities present in minute quantities, probably due to the unusually intense ultra violet radiation to which the water flowing along the lamp is exposed. Such a deposit not only causes a lower light transmission and therefore a decrease in the yield of light, but it is also accompanied by a local overheating of the tube wall due to the greater absorption and the poorer transfer of heat to the water. This may result in the attack on the inside of the the tube (etching) by the mercury vapour, which then more quickly shows a tendency to break at the spots attacked.

Closed cooling system

All these objections can be avoided by providing the mercury lamp with a cooling system with a closed water circulation. The danger of freezing

[7]) For special cases other constructions are also used besides the one described in the following. The inlet and outlet of the water may also be at opposite ends of the lamp.

can then be avoided by adding a suitable anti-freeze substance to the water, while the water can be kept free of all impurities which might be subject to decomposition. As anti-freeze, alcohol with a trace of sodium phosphate is usually used. Because of the danger of decomposition it was found necessary to take care that the alcohol is also free of impurities. It was also found necessary to use a greaseless seal for the axle of the circulation pump, since the small amounts of grease which entered the water with the ordinary packing were also decomposed and again caused a deposit on the lamp.

The cooling-system constructed contains only about $1^1/_4$ litres of water. Considering the fact that

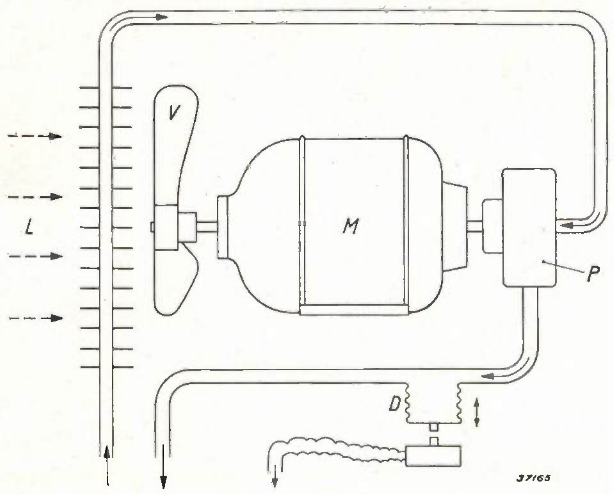


Fig. 8. Construction of the cooling system. The centrifugal pump *P* keeps the water in circulation. The electromotor *M* drives both the pump and the fan *V* which serves for the cooling of the radiator *L*. *D* is a pressure automaton which switches off the current when the water circulation is defective.

for the mercury lamp SP 2 000 W, according to the above, 6 to 8 litres of cooling water per minute are necessary, after leaving the lamp the water has only about 10 sec to lose the heat it has taken up and to be cooled to the initial temperature. For this purpose a radiator of the type used in motor cars has been used. The radiator is cooled by means of a fan on the same axle as the centrifugal pump which makes the water circulate and driven by an electromotor of 150 W. *Fig. 8* and *9* show the arrangement.

If for some reason the circulation stops, then in order to prevent destruction of the mercury lamp the current must immediately be cut off. In the inlet connection of the cooling water to the lamp a pressure automaton is introduced for this purpose; its construction is indicated in fig. 8: the cooling water fills an accordion-like tube which expands more or less according to the water pressure and thereby depresses a push button which operates a switch in the current circuit of the lamp. If the pressure falls

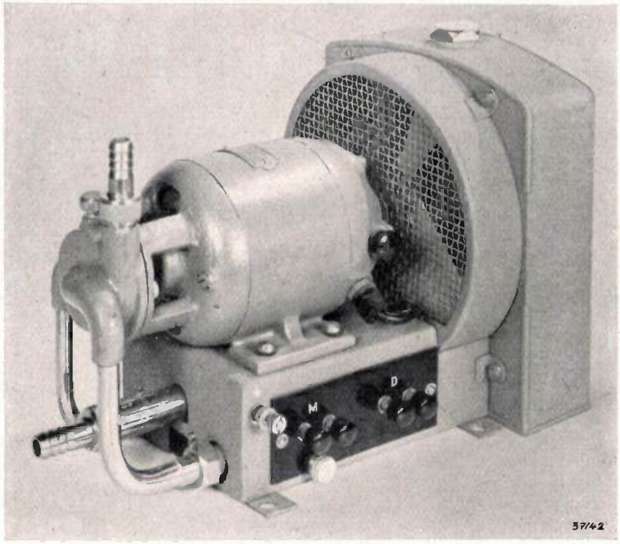


Fig. 9. Photograph of the cooling system.

below a certain value, *i.e.* if the push button comes up far enough, the switch reverses and interrupts the current supply. By means of a setting screw on the push button, the pressure can be regulated at which the switching off will occur.

Fig. 10 shows how the cooling system and the reflector are combined to a unit. In this arrangement the air current sent through the radiator by the fan also contributes to the cooling of the reflector. The chassis is constructed as a section, *i.e.* if desired different complete systems can be set one above the other when one lamp is not sufficient. The reflector can be rotated about its focal line by means of a

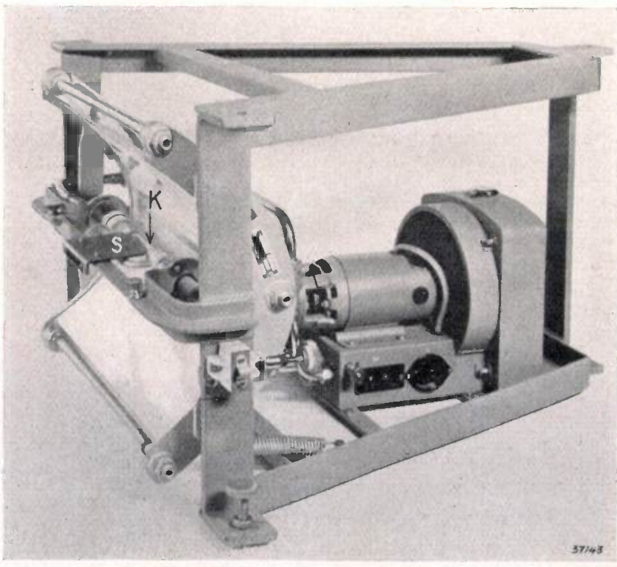


Fig. 10. Chassis with reflector and cooling system. The plate *S* screens off the light from the mercury lamp *K* which is directed upwards in order to prevent blinding. The fan contributes to the cooling of the reflector. Several units like the one shown can be piled one on top of the other.

Fig. 11. Three units one above the other (fig. 10) are here mounted in a rain and sand-proof house.

screw, so that the beam can easily be set at the right angle, or when several lamps are used simultaneously the beams can be made parallel. The systems piled one on top of the other are mounted in a rain and sandproof house with a glass window, see *fig. 11*, so that the mirrors are protected against atmospheric influences. The necessary air for the cooling of the radiator is let in and out through a number of slits behind which a kind of labyrinth is placed to prevent the passage of rain, dust, etc.

For aerodromes in the country which cannot be connected to electric mains, or for cases where no lines can be laid around the whole aerodrome in order to obtain the necessary freedom in setting up the landing light, the latter is combined with a motor aggregate on wheels for generating the current. Such a unit is shown in *fig. 12*.

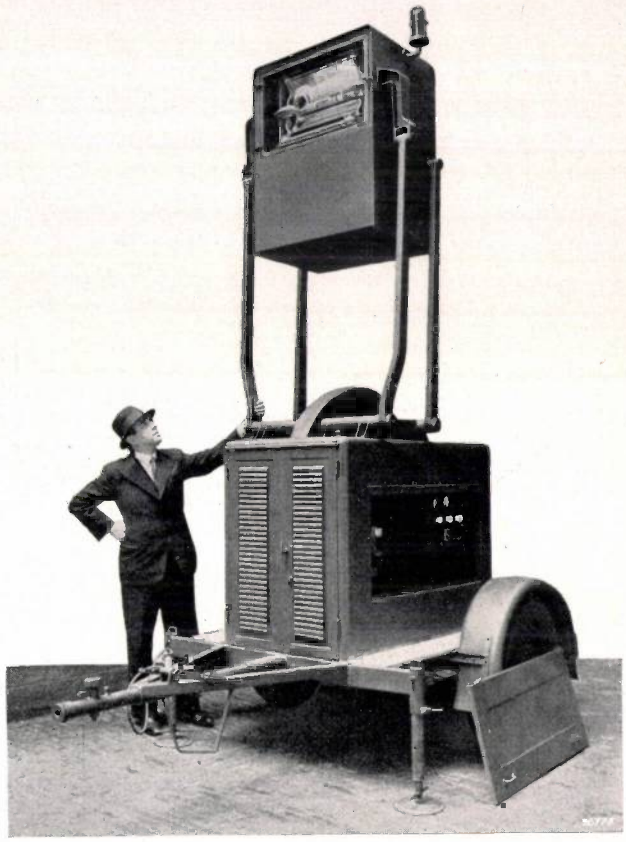


Fig. 12. Combination of a landing light with a petrol engine plus dynamo for the generation of the current. By means of a built-in winch and the lever parallelogram visible on the aggregate the light can be brought to any height between 2 and 4 metres. By adjusting the length of the feet of the car the fan-shaped beam can be made parallel to the surface of the landing area, and then by rotating the reflector about its focal line the beam may be directed on the field at the desired angle.

A RECTIFIER FOR SMALL TELEPHONE EXCHANGES

by H. A. W. KLINKHAMER. 621.314.634 : 621.316.722

For the supply of telephone exchanges which function entirely without supervision a connection in parallel of a rectifier and a battery is ordinarily used. The current-voltage characteristic of the rectifier must in this case, as explained in this article, satisfy very special requirements in order to ensure the desired constancy of the supply voltage, and at the same time to keep the battery in good condition. A description is given of a rectifier developed by Philips, whose characteristic possesses the desired general form. The calculation of the characterisitic shows that it consists of two separate branches, one for normal and one for heavy loading. As appears from a closer examination, this results in a particularly satisfactory behaviour of the rectifier when in use. Moreover, the output voltage of the rectifier is practically independent of mains voltage variations.

For the supply of telephone exchanges a D.C. voltage is in general required which must be able to be kept constant within very narrow limits. At too low a voltage the selecting elements, which make the connections, work too slowly, so that a wrong number may be obtained. At too high a voltage, spark formation, which is very detrimental to the life of the installation, increases very much, and this is intensified still more by the more rapid working of the selectors. In the case of the installations of the Netherlands Post, Telephone and Telegraph the condition is made that the supply voltage of the large exchanges must remain between 60 and 62 volts, independent of variations in the load.

Usually a combination of an accumulator battery and a converter or rectifier is used for the supply. The converter or rectifier, whose output voltage must remain within the limits given, even upon temporary voltage variations of the A.C. mains, serves for ordinary purposes, while the battery which is less economical in use, serves as a reserve when the mains voltage is interrupted by a defect (emergency use). Moreover, the battery serves to combat "cross talk". The microphones of all telephone subscribers are connected in parallel with the voltage source and are therefore coupled *via* the internal resistance of the latter. Current variations in one microphone therefore cause similar variations in all the other microphones unless the coupling resistance is made extremely small. This is accomplished by the permanent connection of the battery with its very low internal resistance, see *fig. 1*. When the battery has had to deliver current once, it must be recharged as quickly as possible in order always to have as large a reserve as possible. Furthermore, in order to keep it in good condition, the battery must be charged several times a month to a slightly higher voltage than the nominal (for example to 66 volts instead of 62 volts).

In the large exchanges the checking of the voltage and the maintenance of the battery can be entrusted to the operating personnel. In the case of small exchanges with only several hundred subscribers, however, one has to do without such personnel; maintaining the constancy of the voltage, in this case within somewhat wider limits, namely 57 to 63 volts, and the care of the battery must here therefore take place automatically. We shall in the following describe a rectifier with selenium valves (type No. 3028) which has been developed by Philips for these unattended exchanges, and which is adapted to a high degree to the special requirements of this purpose.

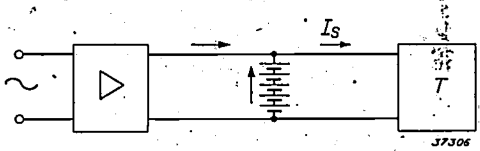


Fig. 1. Connection in parallel of rectifier and battery. The loading current I_s is distributed over selecting elements, microphones, etc. in the telephone network in connection, and the exchange T.

The general nature of the characteristic of the rectifier

It is easily understandable that an ordinary rectifier, *i.e.* a collection of valves, connected to the mains *via* an ordinary transformer, would be useless for the purpose in view. In the first place the D.C. voltage of 60 volts obtained, for instance, might already vary about 12 volts, due to the fact that the mains voltage may rise or fall as much as ten per cent above or below the nominal value. In addition there is still the voltage variation due to changes in the loading. The current-voltage characteristic of an ordinary 60 volt rectifier with blocking-layer valves exhibits a fall in voltage due to the voltage loss in the valves, which with full load may amount to 15 to 20 volts. If these variations are compared with the above-mentioned margin of 6 volts, namely from 63 to 57 volts, it is clear that different methods must be applied in the construction.

The permissible voltage limits for ordinary use actually become still narrower due to the following. The battery voltage is equal to the rectifier voltage, and with voltage variations which are not too rapid, the state of charge of the battery follows this voltage. It may now occur that the mains voltage falls back just at the moment when the battery voltage had adjusted itself at the lowest value for ordinary use. If one is to be sure that the battery in an emergency can then maintain the voltage above 57 volts for a given number of hours, the lowest voltage value for normal use may not lie below, say 60 volts, see *fig. 2.* The voltage margin for the rectifier, *i.e.* the permissible fall in voltage in the ordinarily employed section of its characteristic (BC in *fig. 3*) is thereby limited to a maximum of 3 volts.

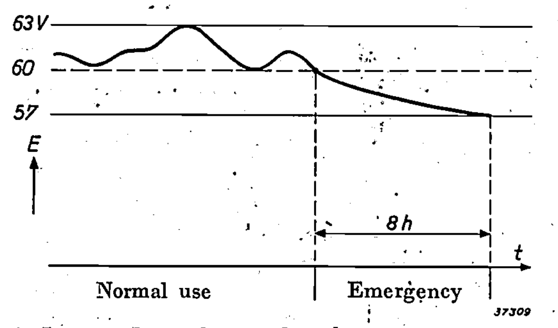


Fig. 2. In normal use the supply voltage may vary between the limits indicated of 63 and 60 volts, for instance. In emergency use, when only the battery functions, it may fall to 57 volts in 8 hours.

In addition to this statement about the part of the characteristic in normal use, it is also possible to say something about the ends of the characteristic, *i.e.* about the behaviour with small and large currents. With large currents a rapid fall is desired in order to prevent overloading of the blocking-layer valves. At very small currents a steep rise of the characteristic is useful (fig. 3). This means that upon very slight consumption the battery will be loaded not only to the normal maximum voltage of 63 volts, but to a still higher voltage, for instance to 66 volts. Since periods of very low consumption regularly occur at week-ends, the above-mentioned requirement of the periodic higher charging of the battery is automatically met. In this way one indeed comes into conflict with the requirement that the telephone automaton connected to the battery may not work at a voltage above 63 volts. Practically, however, no disadvantage is felt when the characteristic in this region is steep enough. A few calls are then already sufficient to cause the voltage to fall to 63 volts again. Upon each call a call converter is switched in which provides the call signal, and which continues to function during the entire conversation. This converter consumes

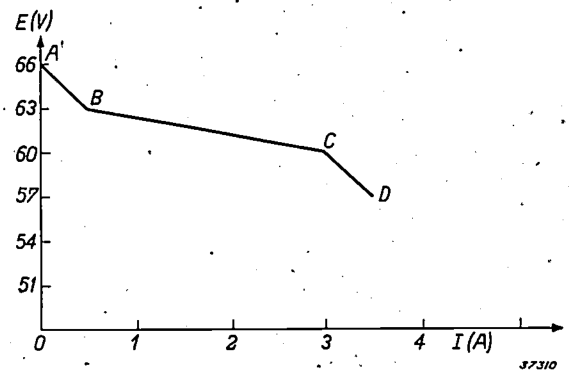


Fig. 3. Desired form of the rectifier characteristic. The section BC is used normally; CD during peak loads, AB in periods of very low loads (week-ends).

a much higher current than is provided by the rectifier at the high voltage of 66 volts. Therefore under these conditions considerable charge is taken from the battery upon every call, while as may be seen from the discharge curve of the battery, only a very small part of its ampere-hours need be taken from the battery to cause its voltage of 66 volts to fall to 63 volts.

The characteristic actually obtained will in the following be compared with fig. 3, which represents roughly the desired form of rectifier characteristic.

Construction of the rectifier

The connections of the rectifier are given in *fig. 4a.* Their action can best be explained by first considering the simpler connections of fig. 4*b* which have long been used by Philips [1]). The iron core of the transformer T_1, which is connected to the mains *via* a condenser, is highly saturated at normal currents, the field of the iron and thus also the secondary voltage E_1 therefore remain practically

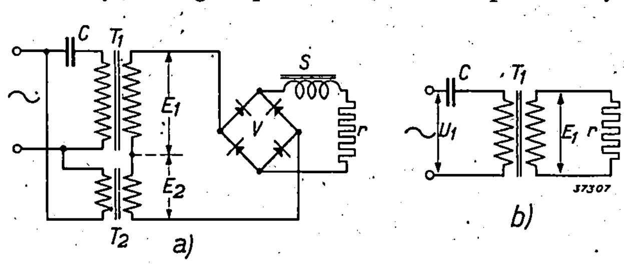


Fig. 4. *a*) Connections of the rectifier. T_2 is a normal transformer, T_1 a transformer with a highly saturated iron core. V are four selenium valves in the Grätz connection, S a smoothing choke, r a loading resistance.
b) Simplified connections. The output voltage E_1 is here insensitive to mains voltage variations, but not yet insensitive to load changes.

[1]) See this periodical **2**, 276, 1937.

constant in spite of fairly large variations of the mains voltage U_1. However, upon variation of the secondary current, as a result, for instance, of a change in the loading resistance in direct connection or connected *via* rectifier valves, E_1 is found to change quite considerably in magnitude and especially in phase (with respect to the mains voltage).

The D.C. voltage obtained by rectification of E_1 is proportional to E_1 in the first instance, and the voltage losses must be subtracted which occur in the valves, in the winding of the transformer, etc. Since E_1 and these losses are all independent of mains voltage fluctuations, this is also true of the D.C. voltage obtained, so that the chief condition is satisfied. Due, however, to the variation of E_1 with the current taken off and to the fact that the voltage losses mentioned increase rapidly with the current, the relation between D.C. voltage and direct current produced is by no means found to possess the desired form. The slope of this characteristic in the ordinary loading region is much steeper than that corresponding to the section BC of fig. 3.

In the complete connections of fig. 4*a* the secondary voltage E_2 of a small transformer T_2 working at normal iron induction is in series with E_1. While the vector E_1 changes in phase and magnitude when the current taken off varies, the vector E_2 thereby remains constant in phase and magnitude. By giving the transformers suitable dimensions, provision may be made that the sum of the two vectors remains nearly constant in the normal region of currents, or even that it increases with increasing current. In this way the characteristic can be given the desired slope. This will become clearer from the calculations to be given presently [2]).

As concerns the behaviour of the rectifier at higher currents than normal, the calculations show that the characteristic here deviates considerably from the form given in fig. 3. The shape obtained however, is found to provide an even more satisfactory functioning of the rectifier than could be expected from the shape of fig. 3.

The calculations are given below.

Calculation of the characteristic

The calculation may be carried out with the help of the equivalent circuit diagram given in *fig. 5a*. The transformer T_1 is here represented by the admittance $-ju$, the valve square with the load connected is replaced by an ohmic resistance with the admittance g. The admittance of the condenser C is assumed to be jv. The transformer T_2 is replaced by the constant A.C. voltage U_2.

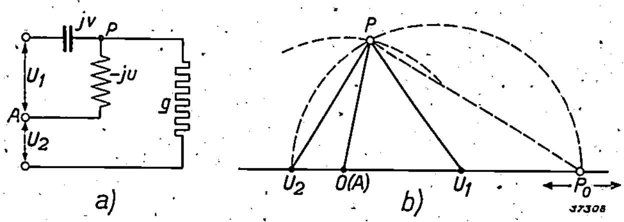


Fig. 5. *a*) Equivalent-circuit diagram for the rectifier connections of fig. 4*a*.
b) Vector diagrams of the equivalent circuit. At a given value of the admittance u (to be considered as a parameter) the point P can be found as the point of intersection of the two circles indicated by broken lines.

At a given value of the admittance g a certain A.C. voltage will act over it and a certain alternating current will flow through it. The relation occurring upon variation of g between these alternating voltages and currents, the "A.C. characteristic", which (except for the voltage losses) gives a picture of the required rectifier characteristic, is constructed by means of the vector diagram. This is of course not permissible, strictly speaking, since the elements of the network are not all independent of the voltage, and the currents and voltages are therefore not sinusoidal: the self-induction u varies when the A.C. voltage acting on it changes, according to the magnetization curve of the transformer core. Nevertheless it is found that no large error is made when the vector diagram valid for sinusoidal voltages is used [3]).

The vector diagram is given in fig. 5*b*. The potential of point A in fig. 5*a* is set equal to zero, the complex potential of point P is indicated by the same letter P. The potentials U_1 and U_2 are opposite because of the direction of winding of the transformer T_2. The voltage P-U_2 acts on g, the current through g is $g(P$-$U_2)$. The behaviour of these two quantities upon variation of g can be derived most simply by using u as a parameter. For every value of g a certain value of u will be effective, and for a given value of u the point P in the diagram may be found as the intersection of two circles: firstly the circle around the origin with the radius $|P|$, which according to the magnetization curve (measured with alternating current) corresponds to the given value of u; P—0 is the A.C. voltage on the self-induction. Secondly the semi-circle constructed through the points which represent the potential U_2 and a potential $P_0 = U_1 \cdot v(v{-}u)$. For the potential P from the condition that the sum of the currents at this node is equal to zero, the following equation is obtained:

$$g\,(P-U_2) = juP + jv(U_1-P),$$

or

$$g\,(P-U_2) = (u-v)\,j\left(P-U_1\,\frac{v}{v-u}\right) \quad . \quad . \quad . \quad (1)$$

Since g and $(u-v)$ are real quantities, the voltage vector

²) Since E_2 varies at the same time as the mains voltage varies, it might be thought that the action of the highly saturated transformer T_1 would to some degree be rendered fruitless. Actually, however, due to the addition of E_2 it is found that the dependence of the output voltage on mains voltage variations can even be still further decreased; see also fig. 10. All this holds only for the normal mains frequency. Upon variations in the frequency the output voltage also varies, relatively about 1.6 times as much as the freqency.

³) The errors can only be large at small output currents ($<$0.7 amp) where the voltage on u and thus the iron induction is the highest. In the following considerations however, it is primarily just with the region of higher currents with which we are concerned.

P-U_2 according to this equation is always perpendicular to the voltage vector P—$U_1 \cdot v(v—u)$, *i.e.* P lies on the semi-circle mentioned. When P has been constructed in this way for a value of u, a point on the required A.C. characteristic is also found, since the distance P—U_2 gives the voltage on g and the distance P—$U_1 \cdot v/(v—u)$ multiplied by $u—v$ according to equation (1) indicates the current through g.

If the construction is carried out for a large number of values of u, then, with the magnetization curve valid in our case, the point P describes the line given in *fig. 6.* From this, by measuring point by point the distances indicated, the A.C. characteristic is found, see *fig. 7.*

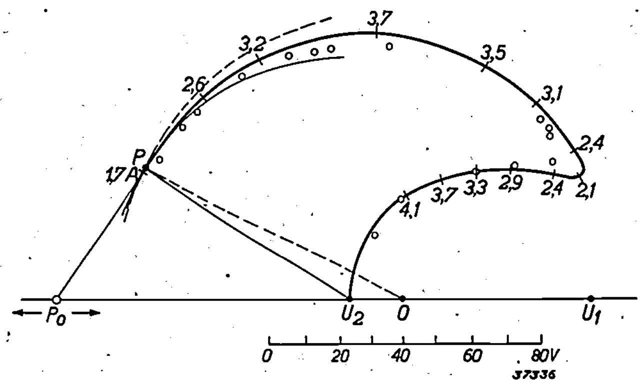


Fig. 6. Line described by the point P found according to fig. 5*b*, when different values are successively chosen for u. At each point P (and corresponding P_0) the length of the two lines P—P_0 and P—U_2 gives two corresponding values of output alternating current (numbers along the curve) and A.C. voltage and thus a point on the "A.C. characteristic". The small circles give several points determined by measurements which are found to coincide well with the calculated curve.

It is instructive to make a comparison between the complete connections of fig. 4*a* and the simplified connections of fig. 4*b*. The vector diagram for the latter would correspond in nature with that in fig. 5*b*, except that U_2 would have to be set equal to zero, and the second circle which furnishes point P would thereby change slightly for each value of u. The curve

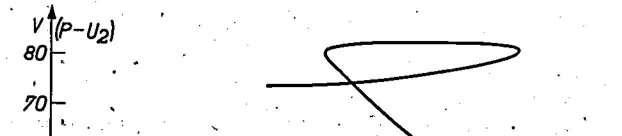


Fig. 7. A.C. characteristic drawn with the help of fig. 6.

in fig. 6 (line described by point P) would in this case remain practically uncharged; as output voltage, however, the distance from P to 0 would have to be taken (broken line), which falls rapidly with increasing current, while on the contrary the distance from P to U_2 occurring as output voltage in the actual connections exhibits an increase with increasing current [4]).

After subtraction of the various voltage losses the D.C. characteristic of the rectifier is found from the curve of fig. 7. Considering the fairly rough approximation, this curve is found to agree quite satisfactorily with the measured characteristic which is reproduced in *fig. 8.*

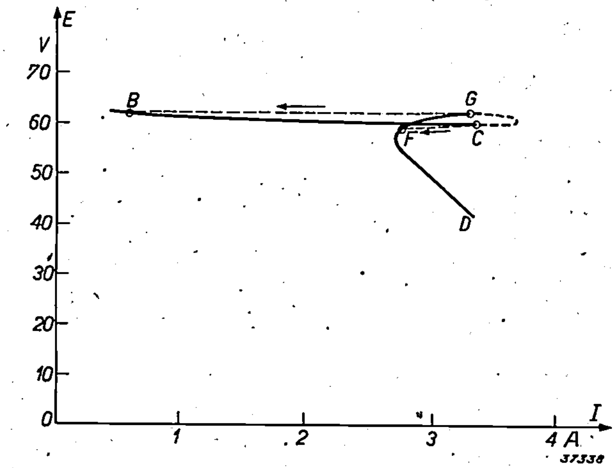


Fig. 8. Measured characteristic of the rectifier. The operating points on the loop which join the branches BC and DFG are not stable, so that a transition between the two branches can only take place by jumps, namely from V to F and from G to B.

Behaviour of the rectifier in action

In the characteristic obtained of fig. 8 it can in the first place be ascertained that the normally used part (corresponding to BC in fig. 3) meets the requirements very well: the output D.C. voltage at a loading with 3 amps. (the highest normally occurring) lies only about 3 volts lower than with $^1/_5$ of this load.

The end of the characteristic toward high currents has a very remarkable shape. A separate branch here occurs which shows a fall similar to that sketched in fig. 3 (CD). This branch is connected by a loop to the rest of the characteristic on which the operating point is situated upon normal loading [5]). Upon closer consideration, however, it is found that not all the operating points on this

[4]) The behaviour of the output voltage can be still better adapted to the requirements if the point U_2 is laid slightly below the extension of the line $0—U_1$. In order to obtain such a phase difference between U_1 and U_2, however, a three-phase mains would be needed.

[5]) Qualitatively it is immediately clear that the connections behave quite differently in this region (low voltages on u): the iron of T_1 is here no longer saturated, the self induction therefore becomes so great, *i.e.* the admittance $-j\omega$ so small, that this branch of the equivalent circuit fig. 5*a* can be entirely neglected. We then have simply a condenser and a resistance connected in series with the voltage $U_1—U_2$; in this region point P in the vector diagram therefore moves along a semi-circle constructed through U_1 and U_2.

loop are stable. When the operating point, moving along the line BC, reaches C where the voltage is 60 volts, it then jumps from there to F on the other branch and runs along the line FD. Moving in the opposite direction, however, it passes from D through F to G, corresponding to a voltage of 63 volts, and then jumps suddenly to B on the first branch.

This remarkable form of the characteristic is particularly favourable for the maintenance of the battery. Every time such a long continued high loading occurs that the battery voltage falls below the limit of 60 volts desired for ordinary use (*i.e.* every time the operating point tries to fall along BC over C), the rectifier begins to work on the branch DFG and continues to work on it even after the conclusion of the peak loading. This means that the battery, discharged to a low voltage, which has had to supplement the rectifier current up to the desired higher value during the peak loading, is again charged with the full rectifier current of 3 to 3.5 amps, until its voltage has risen to 63 volts (operating point G). Only then does the rectifier current suddenly fall to a lower value (0.7 amp. at point B). One may therefore be sure that after a peak loading the whole reserve of the fully charged battery will be available in the shortest possible time, and on an average the battery voltage will be much closer to the maximum of 63 volts than could be reached with a characteristic according to fig. 3.

The behaviour described is well illustrated by the record given in *fig. 9*, on which the current is recorded which was delivered by the Philips rectifier during several hours of normal use in a small so-called end exchange of the Netherlands Telephone. The strong fluctuations in which the current does not rise above about 2 amps, indicate that the rectifier continually adapts its current delivery to the load, and the operating point shifts back and forth on the section between B and C of the characteristic. At several load peaks, however, it may be seen that the current rises to 3 amps; at these points the battery voltage was 60 volts. The current is then maintained for a moment at the high value (the slight variation in these truncated high peaks indicated a shift of the operating point to the branch DFG) until the battery voltage has again reached 63 volts. At the end of each peak the current suddenly falls again to a low value.

Upon variations in the mains voltage the rectifier voltage remains practically unchanged, as was stated above. This is clearly shown in *fig. 10*, where the characteristic measured is given for three values of the mains voltage. It is remarkable that the output voltage of the rectifier for a low mains voltage lies even slightly higher than for a high mains voltage.

If after a long continued defect in the mains the battery is entirely discharged, and it is desired to charge it up as quickly as possible, the rectifier can furnish a higher current of for instance 4 amps by a commutation (T_2 in fig. 4*a* is short circuited, C is increased).

The "week-end voltage"

In the above calculation and discussion of the rectifier characteristic it did not appear whether or to what degree the apparatus described satisfies the requirement that at very small currents the voltage must rise approximately as in the section AB of the curve in fig. 3.

If we now consider more closely the connections consisting of a rectifier, a smoothing choke, a bat-

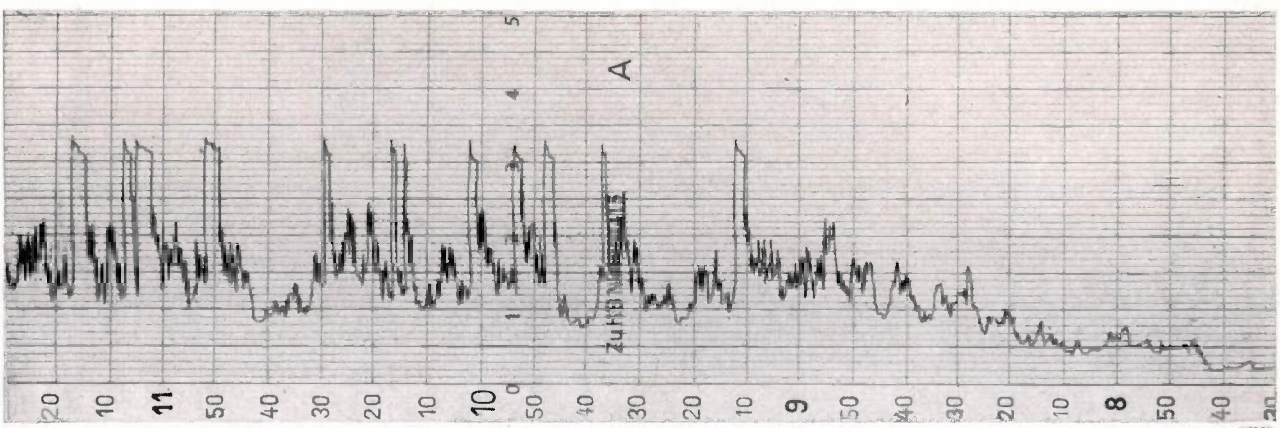


Fig. 9. Registrogram of the direct current delivered by the rectifier during several hours of normal use in a small so-called end exchange of the Netherlands Telephone. At the truncated peaks of about 3 amps. the rectifier was working on the steep branch of the characteristic (DFG in fig. 8). .

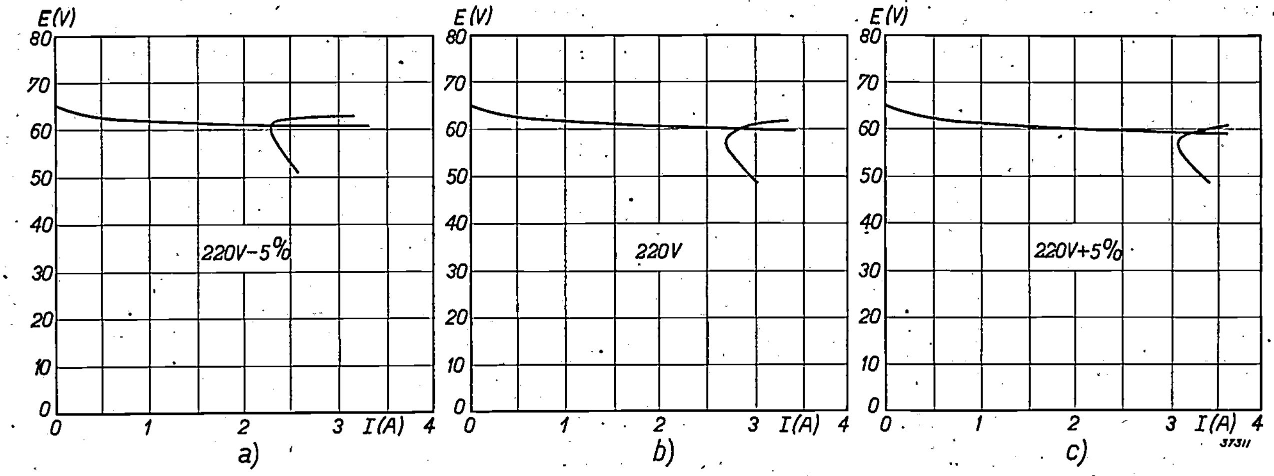


Fig. 10. Measured characteristic of the rectifier at a mains voltage lowered 5% (a), at normal voltage (b), and at a mains voltage raised 5% (c).

tery and a loading resistance (*fig. 11a*), it is found that such a rise in the characteristic as mentioned above occurs of itself. In order to understand this one must first consider the circuit with the battery missing. The rectifier gives a commutated A.C. voltage (fig. 11b), the choke, however, which acts as a kind of "flywheel", keeps the current through the resistance constant except for a slight ripple. Therefore an almost constant voltage acts on the resistance which is equal to the average value of the voltage of fig. 11b. If the battery is now connected in parallel with the resistance and if its voltage is exactly equal to the average value mentioned, the circumstances remain the same. It is true that in the shaded part of the curve the counter EMF of the battery is higher than the voltage given by the rectifier; one might therefore think that the valves would be loaded in the blocking direction and thus no longer transmit current. Actually, however, the choke delivers a voltage which supplements the momentary rectifier voltage to the amount of the counter EMF. In other words, the choke by its flywheel action prevents the current interruptions which might occur. The choke can do this when its self-induction and the average current are high enough. If the latter is now made smaller and smaller, the choke will finally no longer be able to deliver the required voltage during the whole shaded period, current interruptions occur. This means, however, that the rectifier begins to furnish a large proportion of the total current at moments when its momentary voltage is higher than the counter EMF of the battery, *i.e.* the battery is charged and the whole adjusts itself to a higher average D.C. voltage. In the extreme case when the load is zero. the process of charging the battery by the peaks of the voltage of fig. 11b will proceed until the battery voltage has become equal to the maximum value of the commutated A.C. voltage.

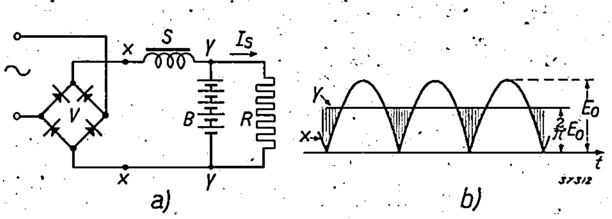


Fig. 11.a) Connections of rectifier valves V_1, choke S, battery B and loading resistance R.
b) Commutated A.C. voltage (peak value E_0), which acts on the terminals x-x in (a). A practically constant voltage with a magnitude of $(2/\pi)$. E_0 acts on the terminals y-y. If the loading current I_s is very small, the rectifier no longer delivers current throughout the whole period of the A.C. voltage, but chiefly at the moments when the commutated A.C. voltage exceeds the counter EMF of the battery.

This is actually what was desired, except for the fact that the rise of the characteristic generally begins at too large currents, since the self-induction of the choke cannot be made indefinitely large, and therefore already at relatively high currents it is no longer sufficient to suppress current interruptions entirely. In order to obtain the desired steepness of the characteristic in this region in spite of this, a second choke with a closed iron core is connected in series with the smoothing choke. At normal currents the iron is saturated, the self-induction is therefore small and the choke may be considered not to be present. At small currents, however, where one still works on the steeply rising branch of the magnetization curve of the iron, the self-induction is very high, so that here the choke still prevents current interruptions down to very small loads. This effect can be seen in *fig. 12*.

If the load remains zero or very small for very long (quiet period of the week-end) the operating point then slowly climbs up the steep left-hand extremity and the battery, as desired, is charged to a voltage higher than 63 volts. The heights of

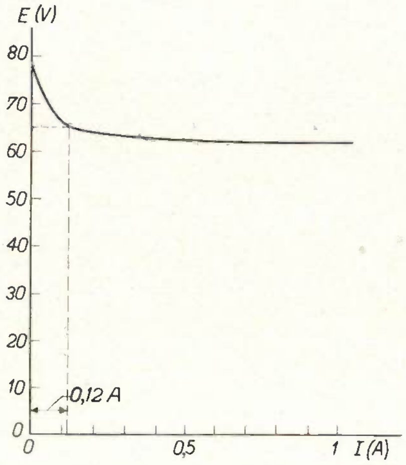


Fig. 12. Beginning of the characteristic of the rectifier with chokes. By means of a small basis load (about 0.12 amp.) such a piece is cut off the left-hand end of the characteristic that the desired maximum voltage ("week-end voltage") occurs.

this maximum voltage will depend upon the circumstances, *i.e.* upon the number and distribution of the calls. In order to ensure that the batteries in all the exchanges are charged to the same excess voltage on Monday morning, a variable resistance (so-called basis load) is connected to the rectifier in parallel with the battery. By taking off in this way a larger or smaller no-load current from the rectifier, a larger or smaller piece of the characteristic is cut off on the left-hand side (fig. 12). Due to the fact that the characteristic rises so steeply here, the "week-end voltage", which without any loading would rise to 75 volts, can be reduced to the desired voltage of for instance 66 volts with a basis load of only a few watts. In the photograph of *fig. 13* the screw with which the basis load and thus the week-end voltage can be regulated may be seen at the upper centre. In this photograph the other components of the circuit may also be seen and are referred to in the text below the figure.

Summarizing it may be said that the rectifier here described, because of its independence of mains voltage variations and because of the form of its characteristic, guarantees satisfactory functioning of the telephone automaton and maintenance of the battery. Thanks to the steepness of the characteristic at large currents, overloading of the blocking-layer valves is made impossible. This also makes possible the direct connection of the rectifier in parallel with a second rectifier, for example upon enlargement of the exchange. The fact that due to unavoidable small differences in the characteristics the load will generally be divided very unevenly over the two apparatus, constitutes no objection because of the impossibility of overloading. The efficiency of the apparatus in normal use amounts to about 65 per cent. The power factor is 0.57, with current leading, however, so that the apparatus helps to improve the power factor of the mains.

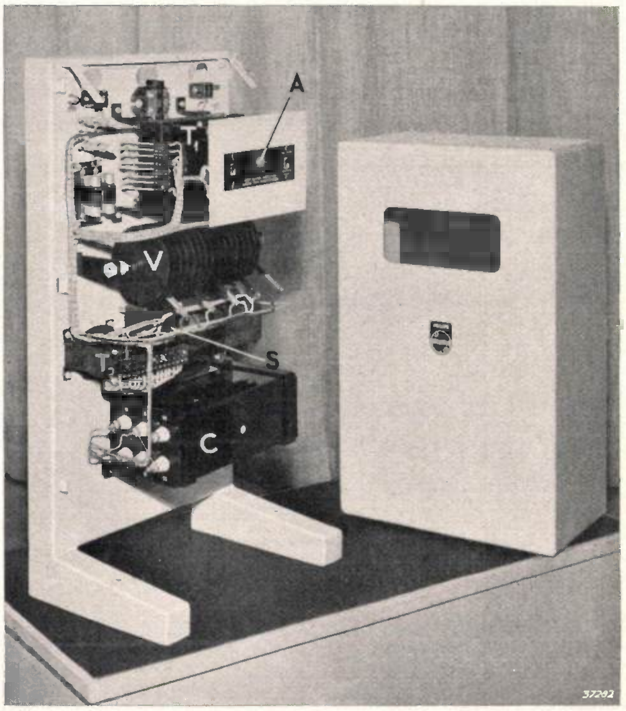


Fig. 13. Photograph of the apparatus opened. *C* condenser (see fig. 4*a*), T_1 highly saturated transformer, T_2 normal transformer, *V* selenium valves, *S* chokes, *A* screw for regulating the "week-end voltage".

AN APPARATUS FOR THE TRANSMUTATION OF ATOMIC NUCLEI

by F. A. HEIJN and A. BOUWERS.

539.172 : 621.319.52

For the transmutation of atomic nuclei use may be made of ions which have been accelerated by a high voltage and with which the atoms to be converted are bombarded. An apparatus for this purpose is described, with which an accelerating voltage of $1^1/_4$ million volts is used, while the substance bombarded is at earth potential. The quantity of neutrons obtainable with this apparatus is approximately the same as could be produced with 5 kilograms of radium. At the end of the article various scientific and medical applications of the installation are discussed.

In recent years there has been a very rapid development of that branch of physics which deals with atomic nuclei. One of the most important pieces of apparatus for research in this subject and for practical applications which have already grown out of the investigations is the apparatus by means of which a very high velocity can be given to certain particles (ions), which are then used to bombard other particles. Two apparatus of this type, which were then called "neutron generators" because of their most important application, have already been described in this periodical [1]). The accelerating voltages were 300 and 600 kV respectively in those cases. In the case of the 300 kV neutron generator the yield of neutrons, although relatively large, was insufficient for many (particularly medical) purposes. The 600 kV apparatus has the disadvantage that the part where the transmutations take place was under high tension and the possibilities of its application were therefore limited. The yield of neutrons of the 600 kV generator was already considerably higher than that of the 300 kV generator, but further enlargement remained desirable.

[1]) F. A. Heijn, The production and use of neutrons, Philips techn. Rev. 3, 331, 1938.

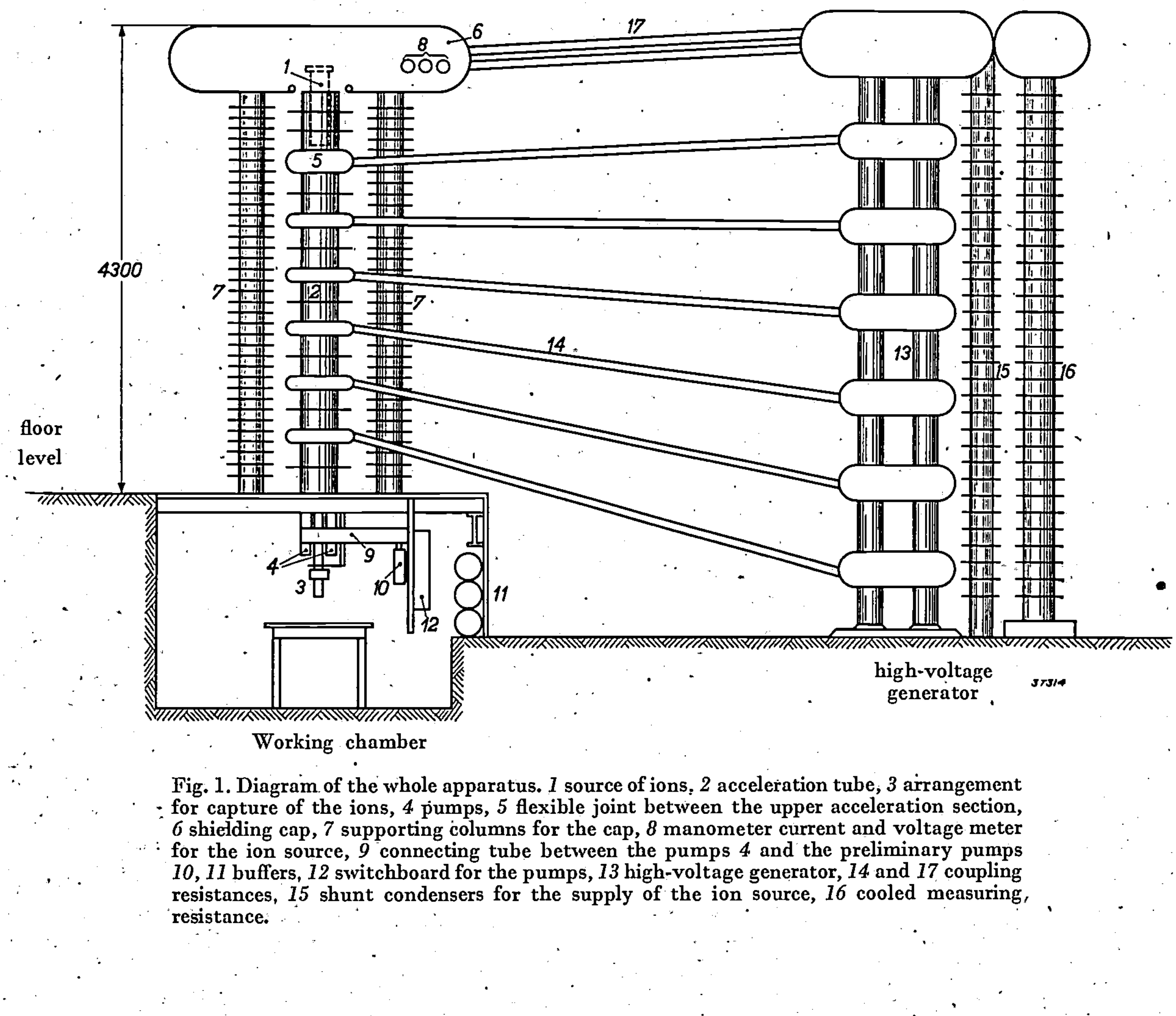


Fig. 1. Diagram of the whole apparatus. *1* source of ions, *2* acceleration tube, *3* arrangement for capture of the ions, *4* pumps, *5* flexible joint between the upper acceleration section, *6* shielding cap, *7* supporting columns for the cap, *8* manometer current and voltage meter for the ion source, *9* connecting tube between the pumps *4* and the preliminary pumps *10*, *11* buffers, *12* switchboard for the pumps, *13* high-voltage generator, *14* and *17* coupling resistances, *15* shunt condensers for the supply of the ion source, *16* cooled measuring resistance.

Therefore a start was made on the construction of a apparatus with an accelerating voltage of 1.25 million volts in which the transmutation chamber is earthed. In the following we shall give a description of this apparatus which is set up in the Philips laboratory, as well as several examples of its application.

Like the previous apparatus, the present one consists mainly of two parts: a tube in which the ions are produced and accelerated by a high voltage, and an installation for generating the necessary high voltage. *Fig. 1* is a diagram of the whole arrangement. Since the available room was not high enough for the high-voltage generator (the generator is about 5.80 m high, while in order to prevent flashover, the ceiling must lie several metres above the top of the generator), it was placed in a pit about 1.30 m deep. The vertically placed tube is placed at floor level; beneath this tube is a chamber 3×3 m excavated to a depth of 2.50 m which can be entered by stairs at the side. In this chamber, in which the investigator can remain with safety while the apparatus is working, the transmutations take place. Entering and leaving the chamber is of course only possible when the high voltage is turned off. The switching desk for the operation of the apparatus is entirely outside the generator room. In this way the person who operates the apparatus is protected against X-rays and other radiation which may be excited at various points, while through several lead glass windows he can still have a good view of the room; moreover, he is in telephonic communication with the investigator in the transmutation chamber.

The tube

In fig. 1 the general construction of the tube may be seen. It contains a source of ions (1), in which the ions used for the transmutation are produced at a potential of 1.25 MV with respect to earth. The ions then "fall", accelerated by the potential difference through the acceleration tube *2*. In the lower section of the tube which projects through the floor into the laboratory below, the ions are captured by the substance which is to be transmuted (at *3*). In this chamber the pump installation (*4*) is also housed, which provides the required low pressure in the tube.

The ion source

It is only necessary to give a brief description of the source of the ions since it is practically the same as that of the previously described apparatus. *Fig. 2* shows a cross section of it. Between the chromium-plated copper anode cylinder A and the aluminium cathode cylinder K a voltage of 50 kV is applied. The whole is filled with the gas whose ions are to be used for the bombardment. In the space S

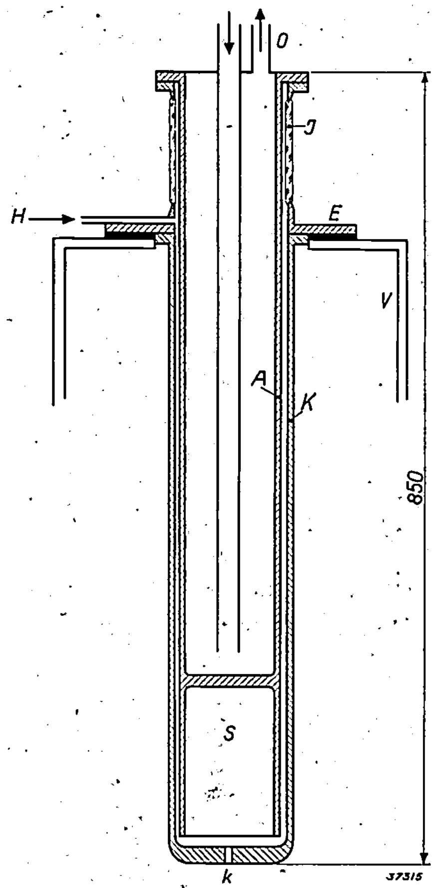


Fig. 2. Cross section of the ion source. A anode, K cathode, S discharge chamber, k channel through which the ions reach the acceleration tube, O lines for the oil cooling, H gas inlet, I insulating glass ring which is fastened by means of welded chrome iron flanges to A and K. With the flange E the ion source is fastened to the acceleration tube V.

a discharge then takes place, and the positive gas ions hereby formed can pass through the channel k drilled in the cathode to the outside, and so into the acceleration tube. Due to the special shape of cathode and anode, at not too high a pressure (0.02 mm Hg), a concentration of the discharge occurs in the neighbourhood of k so that a large percentage of the ions are available for acceleration. The anode must dissipate an energy of several kW; an oil-cooling system is therefore applied (oil lines O), about which we shall say more later.

The ion source, after being completely assembled, is inserted into the top of the acceleration tube with a rubber ring under the flange E. It is unnecessary to screw the flange tight, since after evacuation the air pressure holds the components firmly together; the rubber ring provides for a vacuum-tight seal.

The acceleration tube

The acceleration of the ions takes place in seven steps. The acceleration tube is subdivided into seven sections (see fig. 1) to each of which is applied $^1/_7$ of the total voltage, thus about 180 kV. In each section are two hollow tube-shaped electrodes (see A and B in *fig. 3*). The ions fly with a constant

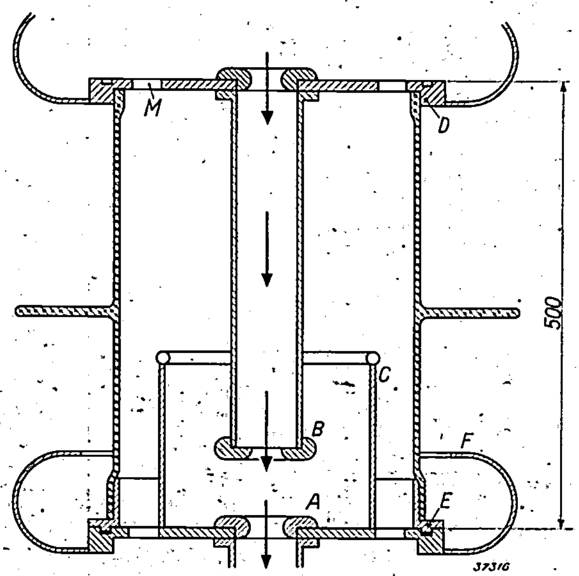


Fig. 3. One of the seven acceleration sections. A and B hollow electrodes through which the ions move and between which they are accelerated with 180 kV. C metal cylinder which surrounds the acceleration chamber. D and E metal flanges for piling the "Philite" cylinders on top of each other, F shielding caps for improving the form of the electric field, M holes in the transverse partitions for decreasing the flow resistance.

velocity through the one hollow electrode, are accelerated in the space between the electrodes with 180 kV, then, again with constant velocity, pass through the following tubular electrode, etc. This subdivision of the voltage is necessary in order to avoid too high a field strength, and thus cold electron emission at certain spots in the tube, while moreover the division of the tube into sections makes it possible everywhere to limit the distances covered by stray ions and electrons to such a degree that they can do no harm.

The sections consist of "Philite" cylinders which are provided with air-tight metal flanges (D and E). The electrodes are fastened to the flanges. The space between the electrodes of each section is further surrounded by a metal cylinder (C). This prevents the "Philite" wall from being struck by electrons which are freed from the cathode of each acceleration section by ions which strike them.

The complete sections are piled on top of each other with rubber rings between for the vacuum seal. Here also it is unnecessary for the sections to be screwed together, since they are pressed together by the air pressure with a force of 750 kg. In this way tubes for any desired acceleration voltage can be built up. Since the high-voltage generator is built on this sectional principle it is possible to begin with a small installation and to enlarge it later on for higher and higher voltages.

The flanges of the sections, set one upon the other, are surrounded by metal caps (F) in order to make the field vary more uniformly and to prevent corona discharge. To these caps the approximately 4 m long resistances are also connected *via* which the voltages are applied to the successive sections. The various details may be seen in the photograph *fig. 4*.

The arrangement for capturing the ions

The arrangement for capturing the ions below the tube is shown in *fig. 5*. After passing through the last acceleration section the ions pass through the tube G, the vacuum tap K and the glass intermediate section H (this and the insulation ring I in the ion source, see fig. 2, is the only glass which occurs in the whole tube), and finally reach the tube T at the bottom of which the actual disc for their capture is situated. This disc consists of the substance which is to be transmuted. Altogether, from the channel k in the ion source to the disc the ions cover a distance of about 4 m. Since in doing this they must pass through a large number of narrow tubes, namely the hollow electrodes, care must be taken that the beam of ions is directed exactly along the axis of the tube. The possibility of making this adjustment is obtained by making the joint between the first and second acceleration sections flexible (5 in fig. 1). By means of a simple mechanism operated from the underground laboratory by an insulating axle, the upper section with the ion source can be given the correct position while the tube is working. The adjustment can be controlled visually by observing the beam of ions on its way through the glass section H. In spite of the high vacuum the beam is visible here as a light streak because of its very great intensity. In order to be able to control the position and the cross section of the beam more accurately, a quartz plate is introduced into section H, which by means of a ground joint can be turned into the beam for a moment. At the spot where it is struck the quartz immediately becomes white-hot, and thus indicates the position of the beam.

Because of the considerable heat development due to the ion bombardment the disc V must be cooled by a rapid stream of water. The same is true of the tap K and other parts of th earrangement for

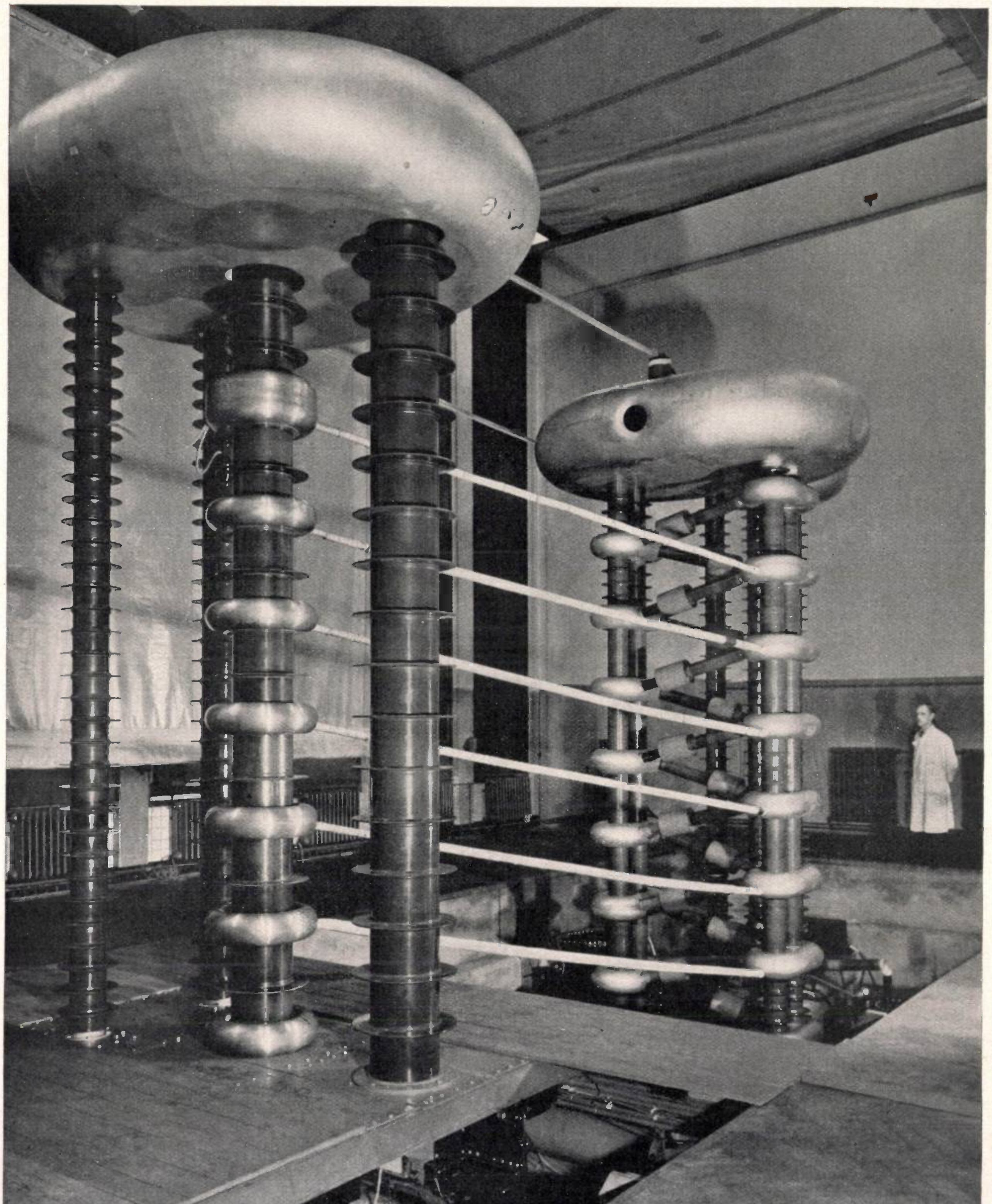


Fig. 4. View of the apparatus. In the foreground the 4 m long acceleration tube with three "Philite" supporting pillars for the shielding cap, and below, through the opening, the laboratory in which the transmutations take place. In the background the high-voltage generator whose successive stages are connected by approximately 4 m long resistances to the corresponding sections of the acceleration tube.

the capture of the ions which might be struck by scattered and reflected ions and thus become too hot (see the cooling spiral L). When the spot on the disc which is bombarded has become useless due to burning in or sputtering the whole arrangement can be shifted slightly, by means of a flexible connection (I) with the acceleration tube, so that the ions strike a different spot. For changing the disc, when for instance another substance is to be transmuted, the acceleration tube can be shut off with the tap K, so that no air enters it when the capturing device is opened.

Fig. 6 is a photograph of the laboratory in which the various details of the arrangement for capturing the ions can be seen.

The pump installation

The ion source only works satisfactorily at not too low pressures (we have already mentioned a pressure of 0.02 mm Hg), while in the acceleration tube the pressure must be as low as possible, preferably lower than 10^{-4} mm Hg. The difference in pressure between the two spaces which are connected by the chnanel k in the cathode of the ion source (fig. 2) can be maintained by vigorous pumping of the acceleration tube; the necessary pumping speed will be lower, the longer and narrower the channel k. This channel is, however, made fairly short and wide (5 mm long, 3 mm diameter) in order to enable the ions to leave the source easily and to obtain a high intensity of the ion beam. Therefore a very high speed of pumping was necessary. In addition, it is only possible to connect the pump to the earthed lower end of the 4 m long tube, and in spite of the great diameter of the tube (30 cm) a not inappreciable loss in speed may occur here. Even with an adequate vacuum at the lower end the pressure in the upper acceleration section would be too high. All measures have been taken to make the flow resistance of the tube as low as possible; for instance the horizontal partitions between the sections are provided with many holes (M in fig. 3). A pumping speed of about 100 l/s at the bottom of the tube was finally found to be necessary. This very speed is obtained by four oil-diffusion pumps in parallel, which together can even reach a speed of 140 l/s (at 10^{-4} mm Hg). A special construction made it possible to connect the four pumps directly to the bottom of the tube (see figs. 5 and 6) so that the loss in speed which any vacuum line would inevitably cause is avoided. The pumps require a fore vacuum of 10^{-3} mm Hg, which is obtained for each two high-vacuum pumps by one oil-diffusion pump of smaller dimensions. These two preliminary pumps in turn require a fore-vacuum (about 0.2 mm Hg) which is provided by one rotating pump of the ordinary construction.

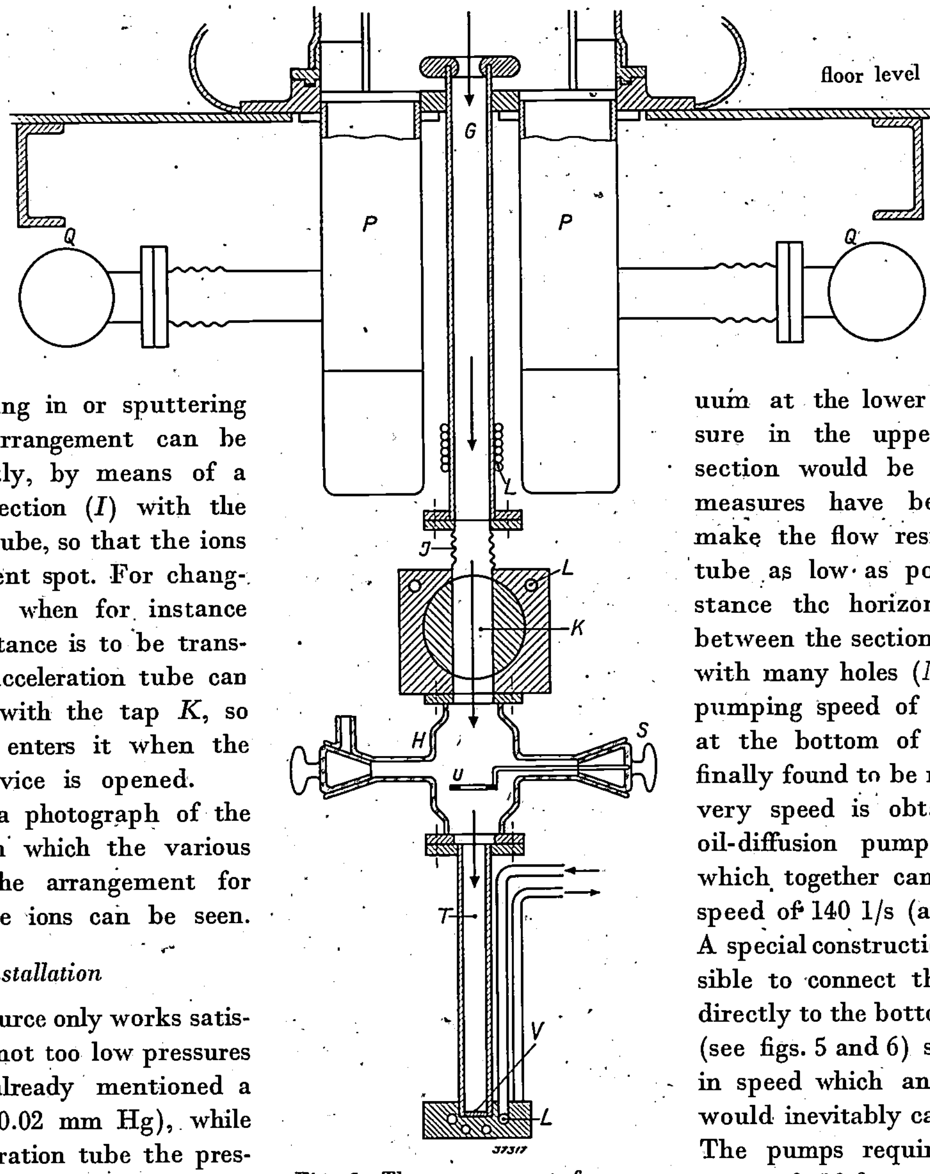


Fig. 5. The arrangement for capturing the ions. G last electrode tube at earth potential, through which the accelerated ions arrive, I flexible joint, K tap, H glass intermediate section, S ground glass joint by which the quartz plate U can be turned into and out of the ion beam, V disc for capturing the ions, L cooling spirals, P pumps, Q connection lines to the preliminary pumps.

The connections (Q in fig. 5) between high-vacuum and fore-vacuum pumps must have a large diameter (10 cm) in order to avoid speed losses here also. Furthermore buffers are introduced between the various pumps in order to ensure a more stable functioning.

The vacuum is continually controlled from the switching table by means of a Philips vacuum meter [2]), while moreover a safety device provides that defects in the pumping system are immediately indicated by lamps on the switching table. This is very important since a lowering of the vacuum may cause a breakdown in the tube which may result in serious damage with the high-voltage used.

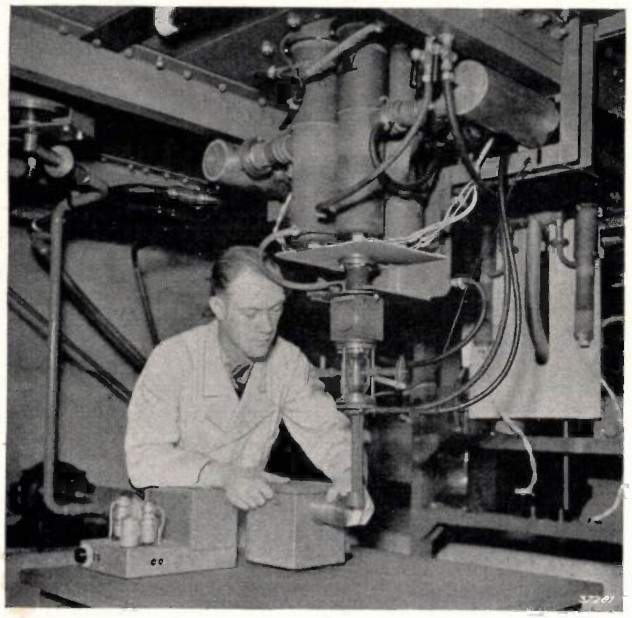


Fig. 6. The laboratory in which the transmutations take place. The glass intermediate section in the arrangement for capturing the ions may be seen, and beneath it the tube at whose end the disc of the substance to be transmuted is placed. Above may be seen the four oil-diffusion pumps grouped around the tube. On the table is a counting apparatus with which for example the intensity of the neutrons radiation obtained can be determined.

Other details

Above the tube is a broad metal shielding cap which is supported by three "Philite" columus (clearly visible in fig. 4), and which fulfils several functions. In the first place it serves to prevent the distortion of the electric field along the tube under the influence of the ceiling, etc. By means of the large radius of curvature of the cap (the smallest radius of curvature occurring is 30 cm) and by smooth finish of the surface the occurrence of of corona discharge is prevented. Furthermore various auxiliary apparatus are housed inside the cap, namely bulbs with supplies of gas for the ion source, a 50 kV rectifier for the supply of the ion source, and parts of the cooling system of the ion source. As to the gas used, most work is done with ions of hydrogen or of deuterium (so-called heavy hydrogen). The necessary gas flows from the bulb in question *via* an adjustable valve into the ion source. According as the pressure in the supply bulb falls, the valve must be opened wider in order to maintain the required pressure in the ion source. Since this adjustment must be able to take place while the apparatus is under high tension, an axle of pertinax for turning the valve is mounted in one of the three "Philite" columns mentioned. This axle is driven by a motor from below which is operated from the switch table. From here also the manometer can be read which is situated in a compartment in the shielding cap (see figs. 1 and 4) and which indicates the pressure in the supply bulb. In two other compartments of the cap are current and voltage meters of the rectifier for the ion source.

The cooling of the anode of the ion source takes place with oil due to the required state of insulation. This oil is let in and out *via* pertinax pipes through the second and third "Philite" columns and *via* a reservoir situated in the cap. Two branches also carry oil to the flange E of the source (see fig. 2) and to the flexible joint in the tube (5 in fig. 1). The oil itself, after its return to earth potential, is cooled by a spiral with water circulation.

It is striking that, apart from the electrical connection and the cooling pipes which are both flexible, the acceleration tube stands entirely free of the shielding cap. This was necessary because, as we have seen, the upper acceleration section must be able to be directed. In spite the small area of the base upon which the 4 m long tube rests, it stands sufficiently firmly because of the air pressure. A rigid mechanical connection with the cap would be undesirable because of the need of being able to assemble the tube easily, and moreover upon vibration of the heavy shielding cap it would be accompanied by the danger of cracks in the tube wall.

The high-voltage installation

The D.C. voltage of $1^1/_4$ million volts used for the acceleration of the ions is obtained with a cascade generator such as has already been described in this periodical [3]). In this case it was built up of seven stages which may be seen in figs. 1 and 4, and whose voltages correspond to the values which are necessary for the seven sections of the acceleration tube. The seven generator stages are connected to the different sections *via* the resistances already mentioned; in this way a potentiometer for 1.25 MV is not needed to distribute the voltage uniformly along the acceleration tube.

[2]) F. M. Penning, Philips techn. Rev. 2, 201, 1937.

[3]) A. Kuntke, A generator for very high direct current voltage, Philips techn. Rev. 2, 161, 1937.

A separate problem is presented by the alternating current supply of the 50 kV rectifier for the ion source at a potential of 1.25 MV with respect to earth. The problem is here solved in the same way as described in the article referred to [1]); the A.C. voltage is applied *via* two condensers (*15* in fig. 1) and two resistances (*17*). These condensers for a *D.C.* voltage of 1.25 MV may well be considered unique. In order to be able to make them of reasonable dimensions an A.C. voltage of 500 c/s was used for the supply instead of the usual 50 c/s; a capacity of 3 000 cm is thereby sufficient.

The source voltage, like the acceleration voltage can easily be regulated at the switching table with this system. The latter voltage is measured with a mA-meter in series with an oil-cooled resistance (*16* in fig. 1). The mA-meter is calibrated directly in kV.

In the photograph fig. 4 a paper screen may be seen above and behind the apparatus. This becomes charged to about half the voltage (600 kV). By this means the field strength at caps and resistances is limited, so that there is less chance of corona discharge and flashover.

Results and examples of application

The apparatus is most commonly used for the production of neutrons [4]). The capture disc for this purpose is made of lithium and is bombarded with deuterium ions, whereby the lithium atoms are converted into two helium atoms and a neutron, according to the reaction equation:

$$\mathrm{Li} + \mathrm{d} \rightarrow 2\,\mathrm{He} + \mathrm{n}$$

The number of neutrons which can be produced with this apparatus amounts to about 10^{11} per second. It is customary to compare this yield with that obtained upon bombarding beryllium with the α-particles of radium. In that case the following reaction takes place:

$$\mathrm{Be} + \alpha \rightarrow \mathrm{C} + \mathrm{n},$$

and 1 mg. radium produces in this way about 20 000 neutrons per second. In order to obtain as many neutrons as with the apparatus described, therefore, about 5 kg radium would be needed! This is much more than exists (purified) in the whole world. For the sake of comparison it may be stated that the previously described 300 and 600 kV installations were equivalent in this respect to about 20 and 300 g radium, respectively.

The neutrons have a similar action upon tissues as X-rays and can therefore, like the latter, be used for therapeutic purposes (such as the cure of cancer) and for all kinds of biological experiments (such as the causing of mutations). Important results have already been obtained in this field. Usually, however, the neutrons are used for the preparation of artificial radioactive substances. Most elements can be converted into radioactive products by irradiation with neutrons. These products are not only of scientific interest, but have also attained great practical importance, again particularly in medicine. For this purpose much radioactive phosphorus has been prepared in recent times with the apparatus described. The following method is used. Around the disc for capture (which is introduced for this purpose into a long narrow tube, *T* in fig. 5) a vessel containing carbon disulphide is placed, while between the disc and the wall of the vessel a voltage of several thousand volts is applied. The neutrons produced on the disc and leaving the tube in all directions transmute the sulphur atoms according to the equation:

$$\mathrm{S} + \mathrm{n} \rightarrow \mathrm{P} + \mathrm{H},$$

whereby the phosphorus atoms formed are radioactive with a half-value time (*i.e.* the time in which the radioactivity has fallen to one half) of 14 days. Moreover these phosphorus atoms are ionized upon formation, and are therefore drawn to the wall by the electric field in the carbon disulphide. Here the phosphorus is deposited in the form of phosphates by a chemical treatment, and it can then be sent to the various medical and other institutions. In this way a number of institutions in the Netherlands as well as in Belgium, Denmark and Italy are provided with radioactive phosphorus. The samples obtained have already been succesfully used for the treatment of leuchaemia patients. The radioactive phosphorus behaves chemically exactly like ordinary phosphorus, and like the latter, when injected subcutaneously, it goes chiefly to the bone marrow, *i.e.* the seat of the disease in leuchaemia [5]).

Due to their property of behaving chemically exactly the same as the corresponding non-radioactive elements, the activated substances are also very suitable as indicators in chemical and biolo-

[4]) On the subject of neutrons in general see the article referred to in footnote [1]) and also, W. de Groot, Nuclear physics, Philips techn. Rev. 2, 97, 1937.

[5]) It may here be noted that the relatively short half-value time of the artificial radioactive substances is essential for the medical application described. Upon injection of radium-containing substances, for example, whose activity practically does not decrease due to the long half-value time (1550 years in the case of radium), the patient would be exposed to a permanent irradiation which in the end would also destroy the healthy tissue.

gical experiments. One then has as it were a number of "tagged" atoms (*i.e.* detectable by their radioactivity), whose adventures among large quantities of the same element or compounds of it can be continually followed. If for exemple the end of a gut twig of some plant is placed in a phosphate solution in which there is also a trace of radioactive phosphorus, several hours later the exchange of phosphorus between the plant and the solution can be demonstrated by the fact that the leaves on the twig have become radioactive! This method is also suitable for quantitative investigations, as may be seen from numerous publications [6]).

In addition to all these application in which the transmutations brought about with the apparatus are only a means to an end, the apparatus also makes possible fruitful investigations of the transmutation itself. Lithium may be bombarded with ordinary hydrogen (instead of heavy hydrogen) and X-rays of great hardness are excited. In order to obtain these rays in the ordinary way with an X-ray tube, one would need an X-ray tube for 17 million volts (which does not exist of course), while for the hardest γ-rays of radium 2.5 million would "already" be sufficient. Such hard rays have very remarkable properties; for example they are not absorbed in matter like ordinary X-rays and γ-rays by giving off energy to electrons and atoms, but by so-called materialization, in which each quantum of the radiation passes over into a positive and a negative electron.

One very remarkable transmutation which was investigated with the apparatus [7]) is the fision of uranium and thorium. These elements fall apart upon bombardment with neutrons and enormous quantities of energy are liberated (corresponding to the energy which an electron would have after passing through a potential difference of 100 million volts). The fragments which are hereby formed could be separated and examined after the bombardment, thanks to the high concentration in which they were formed with this apparatus: they were found to consist of the rare gases xenon and krypton, among other products, and these in a radioactive form. By further fision ceasium, rubidium and barium are formed from them.

It is true that the amounts of the transmuted substances obtained are still extremely small from the chemical point of view and only in exceptional cases could they be detected with a balance. Nevertheless it may be stated that at least in principle the dream of the alchemists has come true.

[6]) See for example: A. H. W. Aten, Isotopes and the formation of milk and eggs, Diss. Utrecht 1939.

[7]) A. H. W. Aten, C. J. Bakker and F. A. Heijn, Nature London, **143**, 516 and 679, 1939.

A DIRECT CURRENT SUPPLY APPARATUS WITH STABILIZED VOLTAGE

by H. J. LINDENHOVIUS and H. RINIA. 621.396.682 : 621.316.722

H.T. power pack supplied by mains or "battery eliminator"

By means of amplifier valves it is possible to stabilize very effectively the output voltage of a plate voltage apparatus, so that it is scarcely affected by variations of the mains voltage or of the current taken off. A direct current supply apparatus with stabilized voltage (GM 4 560) developed by Philips is described. The output voltage of this apparatus is variable between 150 and 300 volts; the current taken off may amount to 100 mA. With the mains voltage fluctuations occurring the output voltage varies less than 0.004 per cent of the value chosen. The internal resistance under most conditions is less than 1 ohm, and under all conditions less than 4 ohms.

1924 !! Is that "recent"?

For many purposes in factories and laboratories a source of direct current is required with a constant terminals voltage which is independent of the current taken from the source. An example of a current source which is capable of satisfying these requirements is formed by the accumulator battery which is indeed used with success in many cases as a source of constant voltage. In recent times, however, a more and more serious attempt may be observed to replace the accumulator batteries by the much lighter plate voltage apparatus which requires much less maintenance. This is true particularly when the source of current must have a voltage of several hundred volts and at the same time need give only a relatively small current. If it is desired to realize such a current source by means of an accumulator battery, a large number of cells must be connected in series, so that even with fairly small cells a large and heavy combination is obtained, while, moreover, such a battery has a relatively large internal resistance and therefore only approximately satisfies the requirement that the voltage shall be independent of the current taken off.

!! If instead of the accumulator battery the familiar plate-voltage apparatus is used, consisting of transformer, rectifier and a filter for smoothing the ripple of the rectified A.C. voltage, the purpose in view is likewise not achieved. Not only does this apparatus possess a considerable internal resistance so that the output voltage is quite appreciably dependent on the current taken off, but also the output voltage fluctuates in the same relation as the voltage of the mains with which the apparatus is connected. Mains voltage fluctuations of 5 per cent are not in the least unusual so that there can be no question of a constant output voltage.

This objection can be met by connecting one or more neon stabilizer tubes between the output terminals of the plate-voltage apparatus. The output voltage is then indeed much less closely dependent upon the mains voltage; the internal resistance, which in this case is practically equal to the differential resistance of the stabilizer tubes, still, however, amounts to from 50 to several hundred ohms.

In this article we shall discuss several automatically working regulatory arrangements which make it possible to keep the output voltage of a plate-voltage apparatus constant within very narrow limits. It will be found that with relatively simple means it is possible to eliminate entirely both the voltage variations due to the changes in the current taken off and those caused by fluctuations in the mains voltage. A direct current supply apparatus with stabilized voltage which has been constructed will be described, which, apart from the perfect constancy of the voltage, also has the advantage over the accumulator battery that its voltage can be varied continuously.

The regulation principle

In principle a constant voltage can be obtained by including in series with an ordinary supply apparatus, for example, a non-regulated plate-voltage apparatus, a resistance which automatically becomes larger when the output voltage has the tendency to increase due to an increase in the mains voltage or to a decrease in the current. With the opposite tendency the resistance must become smaller. This can be obtained by using as resistance a triode whose control-grid voltage is made dependent in a suitable manner on the difference between the output voltage V_u and a constant comparison voltage V_b. This comparison voltage may for example be obtained with the help of a dry battery. If the connections are so arranged that the source need not furnish any current for the comparison voltage, it is not difficult to keep this voltage constant.

A simple example of such a regulatory connection is indicated in *fig. 1*. An ordinary plate-voltage apparatus is connected to the left-hand terminals. the voltage is taken from the right-hand terminals. The grid of the triode is connected to the positive pole of a dry battery, whose negative pole is connected to the negative output terminal.

The output voltage V_u will adjust itself at a slightly higher value than that of the battery, and

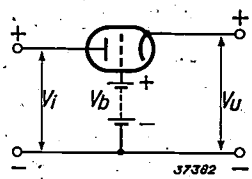


Fig. 1. Circuit diagram showing the principle of the stabilization of a voltage. If the input voltage V_i rises while the current given off is kept constant, the output voltage V_u rises only slightly since the increase of the negative grid voltage occurring as a result need only be small to compensate the influence of the plate voltage increase which results from the input voltage increase. If the current given off i increases while the input voltage is left unchanged, the output voltage falls only slightly, since the decrease of the negative grid voltage occurring as a result need only be small to increase the current i by the desired amount.

in such a way that the grid of the triode possesses exactly the negative voltage with respect to the cathode which is necessary to transmit the current i to be furnished externally. If we now increase the input voltage V_i and thus the plate voltage of the triode, while we leave the current given off unaltered, then the voltage V_u will also increase slightly, and to such a degree that the increase of the negative grid voltage which occurs as a result just compensates the effect of the plate-voltage increase on the triode current. Since a slight change in grid voltage is able to nullify the effect of a large change in plate voltage, the variation in output voltage caused is only a fraction of the variation of the input voltage. In the following we shall call the following ratio:

$$\left(\frac{\Delta V_u}{\Delta V_i}\right)_{\Delta i = 0}$$

the attenuation factor a.

If we increase the current given off while leaving the input voltage unchanged, the output voltage will fall somewhat; as a result the grid voltage of the triode becomes somewhat less negative, as much as is necessary to increase the triode current by the desired amount.

The ratio

$$\left(-\frac{\Delta V_u}{\Delta i}\right)_{\Delta V_i = 0}$$

is called the internal resistance R_i of the regulatory connections, as usual.

We now wish to determine a and R_i quantitatively and we begin with the familiar relation

$$\Delta i = S\,\Delta V_g + \frac{S}{\mu}\,\Delta V_a, \quad \ldots \quad (1)$$

which gives the relation between variations in the grid voltage V_g, the plate voltage V_a and the plate current i, and in which S is the slope and μ the amplification factor. For our case i is also the current given off and the following is true:

$\Delta V_g = -\Delta V_u$, $\Delta V_a = \Delta V_i - \Delta V_u$, and therefore

$$\Delta i = -S\left(1 + \frac{1}{\mu}\right)\Delta V_u + \frac{S}{\mu}\,\Delta V_i. \quad \ldots \quad (1a)$$

From this it is easily derived that:

$$a = \left(\frac{\Delta V_u}{\Delta V_i}\right)_{\Delta i = 0} = \frac{1}{\mu + 1} \quad \ldots \quad (2)$$

and
$$R_i = \left(-\frac{\Delta V_u}{\Delta_i}\right)_{\Delta V_i = 0} = \frac{\mu}{S\,(\mu+1)} \quad (3)$$

By approximation therefore $a = 1/\mu$ and $R_i = 1/S$.

From (1a) follows also the attenuation which occurs for the case when the apparatus is loaded with a constant external resistance R_u. Then instead of $\Delta i = 0$ we have $\Delta i = \Delta V_u/R_u$. The result is:

$$\frac{\Delta V_u}{\Delta V_i} = \frac{1}{\mu + 1 + \dfrac{\mu}{R_u S}}$$

We must also discover what value is assumed by the internal resistance in the case when the supply apparatus itself which produces the voltage V_i possesses an internal resistance of R_v; then ΔV_i is not equal to zero, but $\Delta V_i = -R_v \Delta_i$ and we find that:

$$R_{\text{total}} = \frac{\mu}{S\,(\mu + 1)} + \frac{R_v}{\mu + 1} \quad \ldots \quad (4)$$

If we compare (4) with (3) we see that the internal resistance is increased by an amount $R_v/(\mu + 1)$. This is clear since an input voltage change of $R_v \Delta i$ according to (2) leads to an output voltage change of $R_v \Delta i/(\mu + 1)$.

In order to obtain an idea of the usual order of magnitude we may fill in in the above formulae: $\mu = 20$, $S = 10$ mA/V and $R_v = 500$ ohms. Formula (3) then gives for a the value 0.048, and formula (4) for the total internal resistance $95 + 24 = 119$ ohms.

Several possibilities of improvement

Continuous adjustment of the voltage

As we have seen above the output voltage adjusts itself at a value slightly higher than that of the battery. The output voltage may thus be altered by changing the voltage of the battery. This can be done by putting more or less cells into connection. The voltage change can then naturally only occur in steps. For a continuous regulation we could introduce a potentiometer across one or more cells; this means, however, a load on the battery which in the course of time will result in an appreciable fall in voltage.

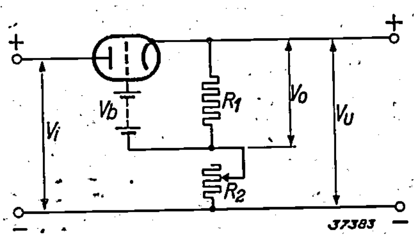


Fig. 2. By not connecting the control grid *via* the battery to the negative output terminal, but to the tap of a potentiometer R_1, R_2, the output voltage is made continuously variable.

In *fig. 2* connections are given in which it is possible, without loading the battery, to regulate the voltage continuously. The negative pole of the battery is not connected with the negative output terminal but to a tap on a variable potentiometer which is introduced between the output terminals of the apparatus. It will be clear that everything which has been derived in the foregoing for the output voltage V_u now holds for the voltage V_0 over the resistance R_1 of the potentiometer. Between V_u and V_0 there is the following relation:

$$V_u = \left(1 + \frac{R_2}{R_1}\right) V_0 .$$

From this it follows in the first place that the output voltage is linearly dependent on the variable resistance R_2. Furthermore it is easy to see that a as well as R_i will be approximately proportional to V_u/V_0. Only the part

$$\frac{R_1}{R_1 + R_2} \Delta V_u = \frac{V_0}{V_u} \Delta V_u$$

of an output voltage variation ΔV_u acts on the grid of the triode. In order to influence the grid in the same way as in the connection of fig. 1, therefore, an output voltage variation which is V_u/V_0 times as large is required. The fluctuations of V_u occurring upon a given change of the input voltage or of the current taken off are therefore found in the first instance to increase by a factor V_u/V_0. From this we see, and this is also true for the following connection, that it is important not to choose the voltage V_0 and thus the battery voltage lower than is necessary in connection with the lowest output voltage which one desires to be able to obtain.

Improvement of the regulation with the help of a second amplifier valve

Although the connections described already lead in certain cases to satisfactory results, for very many applications it is desirable to be able to satisfy still higher requirements. Very considerable improvement can be attained by amplifying the fluctuations of the grid voltage of the regulatory valve with a second valve.

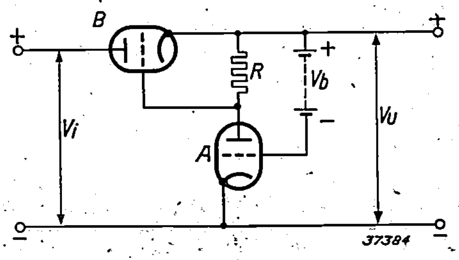


Fig. 3. Stabilizer connections with increased sensitivity. The voltage variations on the control grid of the regulatory valve B are amplified with the help of a second valve A.

In *fig. 3* the connections are given. When the output voltage rises by an amount ΔV_u the grid voltage of value A becomes higher (less negative) by the same amount; the plate current through R thereby rises, and if the amplification is n, the grid voltage of B will become more negative by an amount $n\Delta V_u$. From this it follows that we can immediately apply our previous results to this case if we only multiply the slope and the amplification factor by a factor n. By analogy with (2) and (4) one thus finds directly:

$$a = \frac{1}{n\,\mu + 1} \approx \frac{1}{n\mu}, \quad \text{(5)}$$

$$R_{\text{total}} = \frac{\mu + R_v S}{S\,(n\mu + 1)} \approx \frac{1}{nS} + \frac{R_v}{n\mu} \quad \text{(6)}$$

The constancy of the output voltage upon fluctuations of the mains voltage as well as upon fluctuations of the current taken off is therefore improved approximately by a factor n.

Application of a neon stabilization tube

There are still several objections to the connections of fig. 3 which will be made clear in the following. The voltage over the resistance R serves as negative grid voltage of B and varies from about 1 to 15 volts, according as the current given off by the apparatus is large or small. The plate current of A therefore also changes in the ratio 1 : 15, which is accompanied by the fact that (as a result of the curvature of the i_a-V_a characteristic) the slope of A, and therefore the amplification n, is very closely dependent upon the current given off by the apparatus.

This objection can be met by including a neon stabilization tube in the connection, as indicated in *fig. 4*. This tube is supplied through a resistance

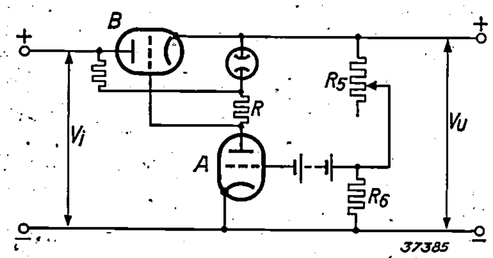


Fig. 4. By not connecting the resistance R of fig. 3 directly, but *via* a neon stabilizer tube, to the positive output terminal of the connections the fall in voltage along this resistance can be increased, and in this way a greater and more constant amplification of valve A can be obtained.

from the voltage source V_i. The resistance R is therefore connected to the positive electrode of the neon tube which has a working voltage of about 100 volts. As we have seen the grid of the valve B has a negative voltage of about 1 to 15 volts with respect to the cathode. It is therefore clear that the voltage over the resistance E in fig. 4 is about 100 volts higher than in fig. 3, and now varies from about 101 to 115 volts. The relative change is thus much less, the amplification n is therefore also much less closely dependent on the current given off by the apparatus. Moreover, the resistance R can be chosen much larger while retaining the same plate current and thus the same slope of A. The amplification will therefore not only be much more constant but also much higher [1]).

Since the amplification factor μ of valve B also depends only slightly upon the current given off, the attenuation factor a, which according to equation (5) amounts to about $1/n\mu$, is practically the same for all values of the current given off. This is true to a much smaller degree of the internal resistance, which according to equation (6) amounts to $1/nS + R_v/n\mu$: the slope of the valve B decreases as the plate current becomes smaller. The result is that the internal resistance becomes larger, the smaller the current given off.

If, in order to form an idea of the results attained, we assume that an amplification $n = 125$ is obtained with valve A, then with the values already used of $\mu = 20$, $S = 10$ mA/V (for valve B) and $R_v = 500$ ohms we find that:

$$a = \frac{1}{2500} = 0{,}4\ {}^0/_{00},$$

$$R_{\text{total}} = 0{,}8 + 0{,}2 = 1{,}0 \text{ ohm}.$$

From this we see that by the use of the second amplifier valve and the stabilization tube we have obtained an improvement by a factor of 120.

The output voltage can be regulated continuously with the help of a potentiometer in a similar way as in the diagram of fig. 2. Here also the attenuation factor and the internal resistance are proportional to the ratio between output voltage and battery voltage. Since no current is taken from the tap of the potentiometer, the resistances R_5 and R_6 of the potentiometer can be chosen as large as desired, so that this potentiometer consumes very little extra energy. In practice, however, this is found to be unnecessary and even undesirable. If only little current is taken from the apparatus, the valve B must be heavily overbiased, and due to this the slope of this valve is very much lowered. It is better not to choose the resistances R_5 and R_6 too large so that the valve B never needs to deliver too small a current. The slope of this valve then remains greater and the values a and R_i then remain smaller and more constant.

Complete elimination of voltage variations

The principle followed in the connections described until now permits indeed a very sharp reduction of the voltage variations, but it is fundamentally impossible to attain a complete elimination of the voltage variations in the manner described. The amplifier valve A, and at the same time the valve B, are controlled by the output voltage; in order to bring the regulatory organ into action it is therefore fundamentally necessary that the output voltage should undergo a certain, although small change. By coupling the control grid of the amplifier valve not only with the output

[1]) It is possible to increase the voltage drop over the resistance R in another way, namely by connecting this resistance directly to the positive input voltage. If the input voltage rises, it is again necessary for good regulation that the grid voltage of B should become more negative. For this purpose, however, a much larger increase in the plate current of A is necessary in this case than in the connections with the neon tube. If in these latter connections we should keep the plate current of A constant, the grid voltage of B would also remain unchanged upon increasing input voltage, while in the other connections with constant plate current of A, the grid potential of B rises as much as the input voltage. The elimination of this rise in the grid potential requires an extra increase in the plate current of A. Further consideration shows that the change in the grid voltage of value A must in this case be a factor μ larger than in the connections with the neon tube. The regulation is therefore also poorer by a factor μ.

side but also with other points in the connections, however, it is possible to compensate the remaining voltage change, and if desired even to attain an over-compensation.

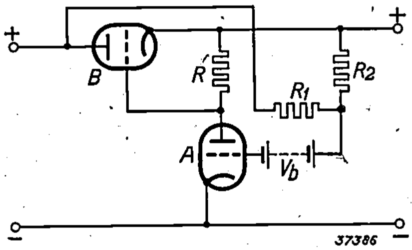


Fig. 5. Connections in which the output voltage remains absolutely constant upon a change in the input voltage.

Fig. 5 shows in what way the influence of a variation of the input voltage on the output voltage can be compensated. If it is desired that the output voltage and the output current flowing through the valve B should remain absolutely constant, while the input voltage varies by an amount ΔV_i, the control grid voltage of valve B must undergo an opposite change by an amount $\Delta V_g = \Delta V_i/\mu$. This change is n times as great as the change in the grid voltage of valve A, so that it must possess the value:

$$\Delta V_{g_A} = \frac{\Delta V_i}{n\mu} \quad \ldots \ldots \ldots (7)$$

Now V_{g_A} is equal to the output voltage, assumed to be absolutely constant, decreased by the likewise absolutely constant battery voltage and the voltage loss over the resistance R_2. The latter varies with the input voltage according to the relation:

$$\Delta V_{R_2} = \Delta V_i \frac{R_2}{R_1 + R_2}.$$

In order to satisfy equation (7) the resistances must be so chosen that

$$\frac{R_2}{R_1 + R_2} = \frac{1}{n\mu} \quad \ldots \ldots \ldots (8)$$

For $n = 125$, $\mu = 20$, this gives $R_1/R_2 = 2\,500$, thus for example $R_1 = 2.5$ megohms, $R_2 = 1\,000$ ohms.

Considering the fact that it is possible in this way to eliminate the voltage variations entirely, it might be asked why it is necessary to make n as large as possible. It would indeed be superfluous to do this if equation (8) could be exactly satisfied for all conditions of working. Since, however, even with the employment of the neon tube according to fig. 4, n as well as μ still vary somewhat with the current given off, this is impossible. Therefore for the majority of conditions of working a certain voltage variation remains which may be positive or negative, and is given approximately by:

$$\frac{\Delta V_u}{\Delta V_i} \approx \frac{1}{n\mu} - \frac{R_2}{R_1 + R_2}.$$

It is thus smaller, the larger the value of n.

If by satisfying equation (8) we have obtained a complete voltage compensation, then the part of the internal resistance which is due to R_v (equations (4) and (6)) also disappears. By a slight extension of the compensation connections, however, the remaining part of the internal resistance can also be exactly compensated. In *fig. 6* the connections

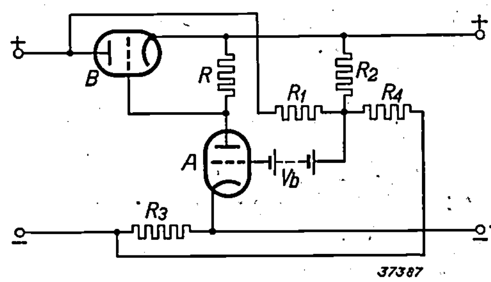


Fig. 6. Connections in which the output voltage remains absolutely constant upon a change in the current given off.

are given. Let us consider a change in current Δi and again assume that the output voltage remains exactly constant. For the change in the control grid voltages of the valves B and A it then follows that

$$\Delta V_{g_B} = \frac{\Delta i}{S}, \text{ where } S \text{ is the slope of value } B, \text{ and}$$

$$\Delta V_{g_A} = \frac{1}{n} \Delta V_{g_B} = \frac{\Delta i}{nS} \quad \ldots \ldots (9)$$

By a correct choice of the resistances R_3 and R_4 the desired change in the control grid voltage of valve A can be obtained. The current variation causes over the resistance R_3 a voltage variation $R_3 \Delta i$, and of this the fraction $R_2/(R_2 + R_4)$ acts upon the control grid of the valve A. (It is hereby assumed that $R_1 \gg R_2$, which is always the case). By setting this equal to the value required according to equation (9), one obtains as condition for the disappearance of the internal resistance the relation

$$R_3 \cdot \frac{R_2}{R_2 + R_4} = \frac{1}{nS} \quad \ldots \ldots (10)$$

With $n = 125$, $S = 10$ mA/V, $R_2 = 1\,000$ ohms and R_4 for instance 50 000 ohms, we obtain $R_3 = 40$ ohms.

Since S changes quite rapidly with the current, and n also to a certain degree, equation (10) will only be exactly valid for a definite value of the current. In general there then remains a certain internal resistance which is given by:

$$R_{\text{total}} = \frac{1}{nS} - \frac{R_2}{R_2 + R_4} R_3 + R_v \left(\frac{1}{n\mu} - \frac{R_2}{R_1 + R_2}\right).$$

A difficulty now arises when we wish to make the voltage variable again. If this were done in the way indicated in *fig. 7* for voltage compensation, the degree of compensation would depend upon the voltage chosen. The part of the input voltage variation ΔV_i which acts on the grid of A is determined by the relation of R_1 to the connection in parallel of R_5 and R_6 [2]), and is therefore dependent

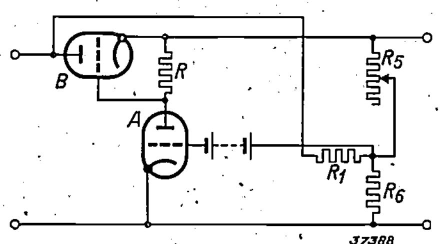


Fig. 7. Connections in which the voltage compensation according to fig. 5 is combined with the variability according to fig. 4. One objection is that the fraction of the input voltage change which acts on the grid of valve A depends upon the value of the resistance R_5 and therefore upon the output voltage chosen.

on R_5. The result would be that when the compensation is correct at low output voltage, there would be over-compensation at high output voltage, *i.e.* there would be a decrease of output voltage upon an increase of input voltage. The same difficulty occurs to the same degree in the compensation of the internal resistance.

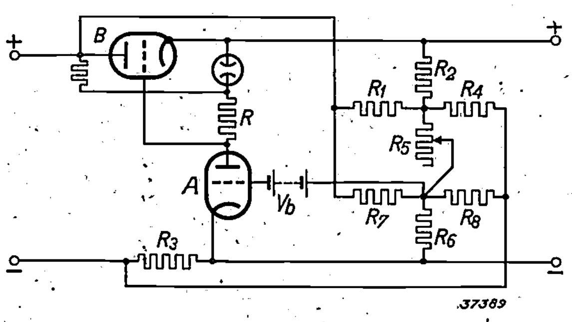


Fig. 8. Connections in which the compensations according to figs. 5 and 6 are combined with the variability according to fig. 4. By means of a bridge connection the disadvantage of the connections of fig. 7 is eliminated.

It is possible, however, with the help of a bridge connection to avoid these difficulties. In the diagram of the complete connections *fig. 8*, this is shown for voltage compensation as well as for compensation of the internal resistance. The bridge connections from this figure are given separately in *fig. 9* for the sake of clearness.

The resistances R_1, R_2, R_6 and R_7 form a Wheatstone bridge in which R_5 is the diagonal. It is

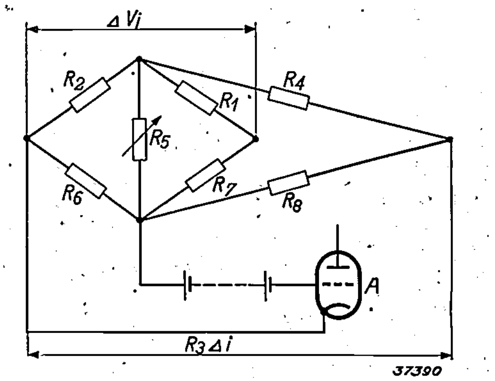


Fig. 9. Diagram of the bridge connections of fig. 8.

immediately clear that the fraction of ΔV_i which acts on the grid of valve A is independent of the resistance R_5 if the following condition is satisfied:

$$R_1 : R_2 = R_7 : R_6.$$

If, in addition, in agreement with equation (8), we make this ratio equal to $n\mu - 1$; the condition for compensation is satisfied at the same time. We have thus succeeded in obtaining the correct voltage compensation for all values of the voltage chosen independent of the magnitude of the resistance R_5.

The same holds for the compensation of the internal resistance. Here the bridge is formed by the resistance R_4, R_2, R_6 and R_8, with R_5 again as diagonal. If again

$$R_4 : R_2 = R_8 : R_6$$

and moreover if, according to equation (10), this ratio is equal to $nSR_3 - 1$, then the compensation of the internal resistance is also correct for all values of the voltage chosen.

The internal resistance for alternating currents

It will often occur that the apparatus, besides direct current, will have to deliver alternating current as well, for instance when it is used as source of supply for an alternating current amplifier. In that case it may be of great importance (for example in order to avoid undesired couplings in amplifiers) that the A.C. voltages thereby occurring at the output terminals should be small; in other words the "internal impedance" for alternating current should be low.

How does this "internal impedance" occur? We have seen that the internal resistance depends

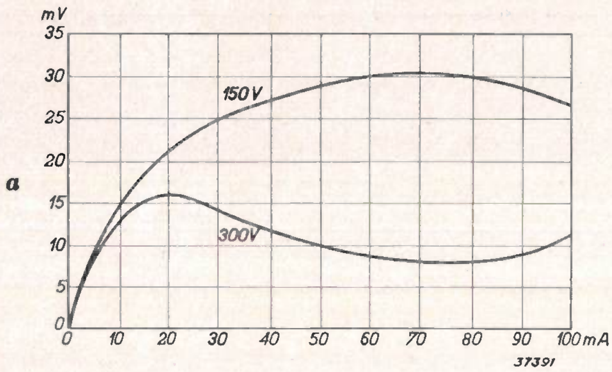
[2]) Since we assume that no voltage variations occur between output terminals, in the investigation of the behaviour of the connections with respect to voltage variations we may consider these terminals as mutually connected.

upon the amplification n of valve A. With increasing frequency this will now decrease due to the capacity C_0 which inevitably acts over the resistance R. The result will be that the internal resistance increases with higher frequencies. It is as if, in series with the internal resistance for direct current R_i, there were an "internal self-induction" L_i. A simple calculation shows that

$$L_i = \frac{V_u}{V_0} \cdot \frac{RC_0}{nS}.$$

The order of magnitude of L_i is several microhenries.

In order to prevent the internal impedance of the apparatus from becoming too high at high frequencies, electrolytic condensers are placed across the output terminals. Then for the case where the internal resistance for direct current is reduced to zero in the way described, the internal impedance consists of a loss-free self-induction in parallel with the electrolytic condensers. If we plot the internal impedance against the frequency, we obtain a resonance curve which, however, is found to be very much damped because of the series resistance of the condensers. By the choice of the capacity and the series resistance we are still able to influence the shape of the resonance curve (the frequency characteristic) to a large extent.



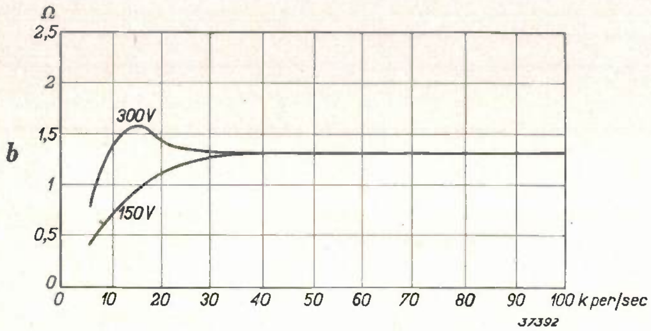


Fig. 10. Diagrams of the direct current supply apparatus GM 4 560.
a) Change in the output voltage as a function of the current given off.
b) Internal impedance as a function of the frequency.

Construction of the apparatus

In the apparatus as constructed (type GM 4 560) the output voltage may be varied continuously between 150 and 300 volts. As comparison voltage

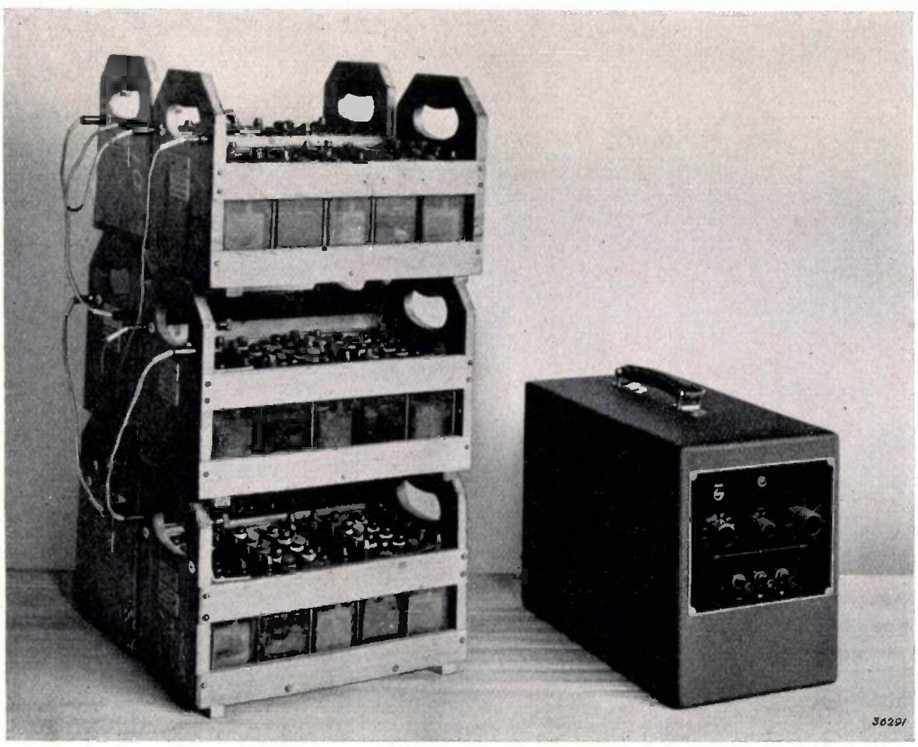


Fig. 11. The direct current supply apparatus compared with an accumulator battery of the same voltage and corresponding output.

a dry battery of 120 volts is used. The maximum current is 100 mA.

A mains voltage variation of 5 per cent causes an output voltage variation of less than 0.004 per cent of the value chosen. The change in the output voltage under the influence of the load is shown in *fig. 10a* for two different output voltages. The slope of this characteristic gives the internal resistance. For currents greater than 30 mA this is found to be a maximum of 0.5 ohm at an adjustment on 150 volts, and 1 ohm maximum at an adjustment on 300 volts. For currents smaller than 30 mA the internal resistance may amount to four times these values.

In fig. 10*b* it is shown for two different values of the output voltage how the internal impedance of the apparatus depends upon the frequency. It may be seen that it is smaller than 1.6 ohms for all frequencies. The fact that at 100 kc/s the impedance still amounts to more than 1 ohm may be ascribed to the series resistance of the electrolytic condensers. At still higher frequencies the impedance is smaller due to the fact that in parallel with the electrolytic condensers a paper condenser is introduced whose series resistance is many times smaller.

The dimensions of the whole apparatus, which is composed of a plate voltage apparatus and the regulatory organ described, are $20 \times 43 \times 31$ cm. The weight is about 19 kg. In this weight is included that of the battery (2.5 kg) which will hardly ever need to be renewed in the course of one year. *Fig. 11* shows the exterior of the apparatus; for the sake of comparison an accumulator battery is also shown which gives about the same power at the same voltage. Weight and volume of the battery are about three times as great.

This sounds rude!

THE CONTROL OF CONTACT PRESSURES

by W. WERNER 621.3.066.6

For the satisfactory functioning of switches, contact springs, etc. it is necessary that the contact pressure should lie between certain limits: at too low a pressure the transition resistance becomes too great and the contact spots may become too hot, at too high a pressure a too rapid wearing off may occur upon repeated closing of the contact. For each case the permissible limits of the contact pressure can best be determined experimentally. For example, for a certain type of mains switches which are used in radio receiving sets, it was found that the contact pressure must lie between 60 and 120 g.

In the mass production of these switches the question arose, as to how it is possible to find out rapidly and simply whether the individual product satisfies the requirement given above. For the determination of pressures of springs several methods are known. In the case in question, however, a particularly simple solution could be found which was specially adapted to the methods of mass production, by making use of the switching action of the contact to be tested itself. The principle of the method is represented in *figs. 1* and *2*. In fig. 1*c* is the switch to be tested (diagrammatic), *a* and *b* are two adjustable auxiliary contacts. In the initial state *a* and *c* are closed, and *b* open. When the lower contact strip of *c* is pulled down, there is a conflict between the elasticity of the springs k_1 and k_2 (to which the contact pressures of *c* and *a*, respectively are proportional). If k_1 is much weaker than k_2 then, upon pulling, at a certain moment the contact *c* will open; if pulled

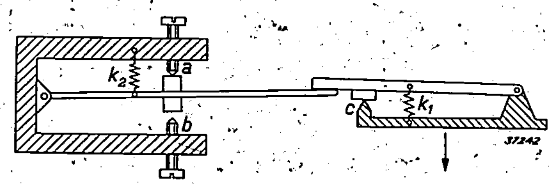


Fig. 1. Principle of the arrangement for testing contact pressures; *c* the switch to be tested; *a* and *b* auxiliary contacts.

farther, at a certain extension of the spring k_1, its force will be large enough to overcome the pull of k_2 so that the contact *a* is then opened. If on the contrary k_2 is much weaker than k_1, then upon pulling, contact *a* will first be opened and only afterwards contact *c*. If k_1 is sufficiently stiff, contact *b* will even first be closed before *c* is opened.

The auxiliary contacts *a* and *b* are so adjusted that *a* is opened when the contact pressure of *c* is just decreased by 60 g by the pulling down, and so that *b* is closed when this decrease amounts to 120 g. Thus if the contact pressure of the switch in question lies within the desired limits, upon pulling down, *a* will first open, then *c* and finally a dry battery of 120 volts is used. The maximum current is 100 mA.

A mains voltage variation of 5 per cent causes an output voltage variation of less than 0.004 per cent of the value chosen. The change in the output voltage under the influence of the load is shown in *fig. 10a* for two different output voltages. The slope of this characteristic gives the internal resistance. For currents greater than 30 mA this is found to be a maximum of 0.5 ohm at an adjustment on 150 volts, and 1 ohm maximum at an adjustment on 300 volts. For currents smaller than 30 mA the internal resistance may amount to four times these values.

In fig. 10*b* it is shown for two different values of the output voltage how the internal impedance of the apparatus depends upon the frequency. It may be seen that it is smaller than 1.6 ohms for all frequencies. The fact that at 100 kc/s the impedance still amounts to more than 1 ohm may be ascribed to the series resistance of the electrolytic condensers. At still higher frequencies the impedance is smaller due to the fact that in parallel with the electrolytic condensers a paper condenser is introduced whose series resistance is many times smaller.

The dimensions of the whole apparatus, which is composed of a plate voltage apparatus and the regulatory organ described, are 20×43×31 cm. The weight is about 19 kg. In this weight is included that of the battery (2.5 kg) which will hardly ever need to be renewed in the course of one year. *Fig. 11* shows the exterior of the apparatus; for the sake of comparison an accumulator battery is also shown which gives about the same power at the same voltage. Weight and volume of the battery are about three times as great.

This sounds rude!

THE CONTROL OF CONTACT PRESSURES

by W. WERNER 621.3.066.6

For the satisfactory functioning of switches, contact springs, etc. it is necessary that the contact pressure should lie between certain limits: at too low a pressure the transition resistance becomes too great and the contact spots may become too hot, at too high a pressure a too rapid wearing off may occur upon repeated closing of the contact. For each case the permissible limits of the contact pressure can best be determined experimentally. For example, for a certain type of mains switches which are used in radio receiving sets, it was found that the contact pressure must lie between 60 and 120 g.

In the mass production of these switches the question arose, as to how it is possible to find out rapidly and simply whether the individual product satisfies the requirement given above. For the determination of pressures of springs several methods are known. In the case in question, however, a particularly simple solution could be found which was specially adapted to the methods of mass production, by making use of the switching action of the contact to be tested itself. The principle of the method is represented in *figs. 1* and *2*. In fig. 1*c* is the switch to be tested (diagrammatic), *a* and *b* are two adjustable auxiliary contacts. In the initial state *a* and *c* are closed, and *b* open. When the lower contact strip of *c* is pulled down, there is a conflict between the elasticity of the springs k_1 and k_2 (to which the contact pressures of *c* and *a*, respectively are proportional). If k_1 is much weaker than k_2 then, upon pulling, at a certain moment the contact *c* will open; if pulled

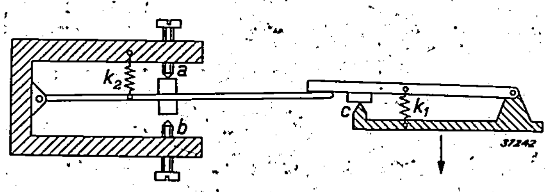


Fig. 1. Principle of the arrangement for testing contact pressures; *c* the switch to be tested; *a* and *b* auxiliary contacts.

farther, at a certain extension of the spring k_1, its force will be large enough to overcome the pull of k_2 so that the contact *a* is then opened. If on the contrary k_2 is much weaker than k_1, then upon pulling, contact *a* will first be opened and only afterwards contact *c*. If k_1 is sufficiently stiff, contact *b* will even first be closed before *c* is opened.

The auxiliary contacts *a* and *b* are so adjusted that *a* is opened when the contact pressure of *c* is just decreased by 60 g by the pulling down, and so that *b* is closed when this decrease amounts to 120 g. Thus if the contact pressure of the switch in question lies within the desired limits, upon pulling down, *a* will first open, then *c* and finally *b* will be closed (order: *a* open, *c* open, *b* closed). Three relays *A*, *B* and *C* are operated by the three contacts *a*, *b* and *c*, and are connected in the manner shown in fig. 2. On the basis of this figure one ascertains that with each of the three switching orders mentioned the final result is that one of the three lamps —, 0 or + burns, by which means the qualification of the contact pressure being tested is obtained as "too weak", "good" or "too strong". By bending the contact springs a deviation can then immediately be improved.

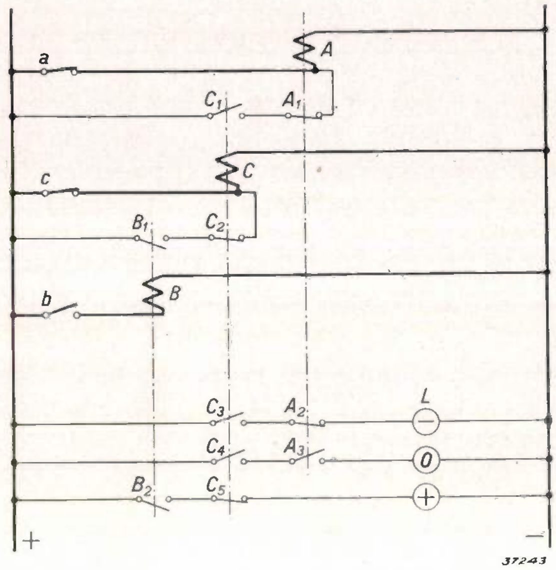


Fig. 2. Connections of the apparatus. The contacts are all shown in the initial state; current hereby flows through the relays *A* and *C* but not through *B*. Upon flow of current through a relay the corresponding switch rod (dash-dot line) is raised. Thus for example if *a* is opened relay *A* opens, with the result that the switches A_1 and A_2 are opened, A_0 closed. Upon opening *c* relay *C* opens, the switches C_1, C_3 and C_4 are thus closed, and C_2 and C_5 opened. Upon closing *b* relay *B* closes and the switches B_1 and B_2 are closed. If the switching order is: *c* open, *a* open, *b* closed, *C* then opens, *A* however remains closed (because C_1 is now closed and *a* therefore shunted), *B* closes; of the three lamps *L* only the one marked- burns. With the switching order: *a* open, *c* open, *b* closed, *A* and *C* open, *B* closes; only lamp O burns. With the order: *a* open, *b* closed, *c* open, *A* opens, *B* closes, *C* remains closed (because B_1 is now closed and *c* is shunted), and lamp + burns.

Fig. 3 is a photograph of the apparatus as constructed. The switch to be tested is slid into the holder *M*, and the upper contact strip is then placed above the pressure finger projecting from the apparatus. Upon pressing the holder down in the manner described one of the three lamps *L* is lighted.

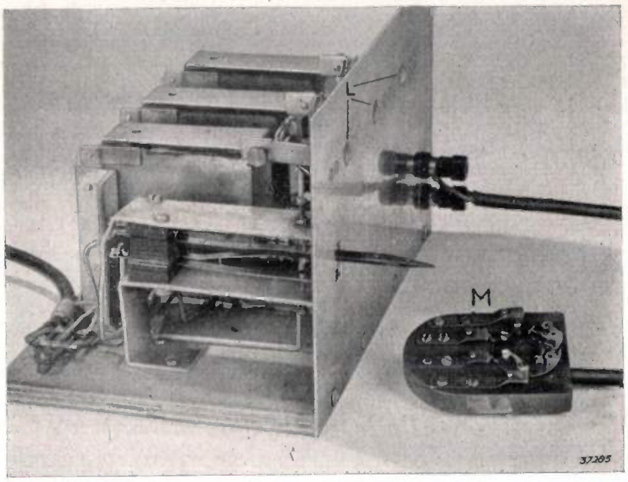


Fig. 3. Photograph of the apparatus. *M* holder, *L* lamps.

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

1509: J. D. Fast: The preparation of metals by powder metallurgy and decomposition metallurgy (Öst. Chem. Z. 43, 27-33 and 48-54, Febr. and Mar. 1940). (Original in German).

Following a consideration of the relationship between powder metallurgy and ceramic processes and of several forerunners of powder metallurgy, the different successive treatments used in the preparation of hard cemented carbides are dealt with and the influence of gases on the sintering process is discussed. Hypotheses are hereby developed about the coming into effect of the forces of cohesion upon compression and about the part which the oxide films always present on the grains of the metal powders play in this process. Furthermore the mechanism of the shrinkage phenomena occurring upon sintering is studied and the phenomena of recrystallization and coalescence occurring at the same time. As an example of the application of powder metallurgy the preparation of ductile tungsten is described, as well as the preparation of the so-called "hard cemented carbides" and several other applications.

Decomposition metallurgy is of particular importance in the preparation in a ductile form of the metals with a high melting point in the titanium group. A description is given of the method used. In conclusion the preparation of pure nickel and iron by the method of decomposition metallurgy is discussed.

The contents of this article have been partially discussed in Philips techn. Rev. 4, 309, 1939 (The preparation of metals in a compact form by pressing and sintering) and 3, 345, 1938 (Zirconium and its compounds with a high melting point).

1510: K. F. Niessen: Calculation of the electrical field strength produced at a given point by a half-wave aerial over a flat earth as a function of the total energy supplied to the aerial I (Physica 7, 586-602, July 1940). (Original in German).

On the basis of an example it is explained how the vertical electric field at a point above a flat earth can be calculated when the field is induced by an aerial of a half wave length, whose lower extremity lies above the earth's surface. It is found possible to express this field with the help of the total energy supplied to the aerial, part of which is radiated (the effective part) and the rest absorbed by the earth (non-effective part). The problem is first solved for the radiation of a pure dipole which in the further treatment is laid at the middle of the aerial. The reflection formula, which is often used as an approximation in such problems, might here lead to quite incorrect results.

1511: J. L. Snoek: On the effective length of a small Barkhausen discontinuity (Physica 7, 609-624, July 1940).

The irreversible part of the magnetization is given by the product of magnitude and number of the so-called Barkhausen discontinuities. At very low field strengths most ferromagnetic substances follow Rayleigh's law in close approximation, according to which law the irreversible part of the magnetization with a small hysteresis loop increases with the square of the maximum field strength. This law furnishes a condition which must be satisfied by the laws which govern the magnitude and number of the Barkhausen discontinuities. Very little is yet known about these laws; it is, however, established that with decreasing field strength the Barkhausen discontinuities become very small. The path of the lines of force of such a very small discontinuity can be calculated in advance, beginning with several plausible assumptions. In this way an "effective length" of the discontinuity is found, and on the basis of this it is possible to choose the most favourable experimental values for the detection of these very small discontinuities. The calculation shows that the effective length is equal to the product of the reversible permeability and the thickness of the wire.

1512: M. J. O. Strutt and K. S. Knol: Determinations of the magnetic permeability from resistance measurements on iron wires of varying structure at frequencies of the order of magnitude of 10^8 c/s., in connection with the magnitude of the elementary zones of Weiss (Physica 7, 635-654, July 1940). (Original in German).

The permeability of iron wires at very high frequencies can with certain assumptions be determined from the ratio between A.C. and D.C. resistance. With the apparatus developed by the authors the A.C. resistance of several carefully prepared wires is measured up to frequencies of 3×10^8 c/s. At room temperature the permeability remains nearly constant up to these frequencies. Upon a lowering of the temperature to —183°C, which for these measurements is equivalent to an increase of the frequency to about 10^9 c/s., a decrease in the permeability takes place. The curves obtained give an indication that the permeability at higher frequencies possibly again assumes a constant value. The hypothesis that this behaviour may be ascribed to a very thin non-magnetic film on the surface of the iron wires is not confirmed by experiments with electrolytically deposited films of more than 10^{-4} cm thickness. A satisfactory interpretation is, however, possible in a manner indicated by R. Becker. This interpretation can be further worked out with a simple model proposed by G. Heller; in particular, with this model the order of magnitude of the elementary zones of Weiss can be determined for the different wires. The results obtained are compared with those of other investigators. In conclusion several unexplained problems are discussed and indications given for possible further experiments.

1513: C. J. Bakker: Het periodiek systeem der elementen (The periodic system of the elements, explanation of the plate). (Ned. T. Natuurk. **7**, 305-310, Aug. 1940).

The plate to be explained represents the periodic system of the elements. In each of the compartments of the chart is indicated: atomic number, atomic symbol, atomic weight according to the international table, electron configuration of the neutral atom in the usual notation, as well as the isotope relations graphically. Chemically analogous elements are indicated by striking colours. Although both by its arrangement of data and by its make-up the plate is already very instructive, in the explanation a further critical discussion of the electron configuration, isotope and radioactive properties may be found. The latter are also indicated in the chart as well.

1514*: J. van Niekerk and Miss M. S. C. Bliek: De methodiek der vitamine-D ijking op kuikens (The methods of vitamin D standardization on chicks) (Landbouwk. T. **52**, 349-353, May, June 1940).

*) An adequite number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on applications to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

This is a communication about a demonstration in Wageningen during the 10th Netherlands Agriculture Week. In the standardization on rats the difference between animalic and vegetable vitamin D is not manifested. This can, however, be established in breeding chicks whereby at the same time the information is obtained whether a product containing vitamin D is suitable for poultry feed and perhaps for cattle feeding also. With a simply composed feed, during the first 3 to 4 weeks of life, rickets occur which can easily be observed in an X-ray photograph on the heel joints of the chicks. Addition of vitamin D in different concentrations to the feed shows a relation between the extent of the rickets and the lack of vitamin D. A preparation consisting entirely of animalic vitamin D gives the antirachitic effect to the same degree as cod liver oil. The relation mentioned between the antirachitic effect and the concentration of animalic vitamin D given is determined in a standardization curve. The exposure technique in the X-ray examination is simple, large groups of chicks can be examined in a few hours.

1515: E. J. W. Verwey and J. H. de Boer: The potential curve of the alkali halide molecules (Rec. Trav. chim. Pays Bas **59**, 633-649, July-Aug. 1940).

To what degree the electrostatic picture of the ionic bond in vapours of the alkali halides is correct in the neighbourhood of the minimum of the potential curve is investigated for three points, namely: 1) the nuclear distances at absolute zero calculated from atomic distances determined with the help of electron diffraction at 1200 °C in the vapour and the atomic distances in the corresponding crystal lattices, 2) the fundamental frequencies of the atomic vibrations, calculated from the potential curves compared with the values from the spectra, and 3) the binding or molecular energies at absolute zero compared with thermal data. The expression for the energy $U = -e^2/r + b/r^n$ with $n = 11$ gives a correct picture. By taking into account the polarizations the picture is improved, and the molecular energies can also be calculated accurately within the errors of observation. For all alkali halides $n = 12$ is then found to be the best value. The exponential repulsion according to Born and Mayer is found incapable of being used.

VOL. 6 No. 3, p. 65-96 MARCH 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

TUBULAR LUMINESCENCE LAMPS FOR GENERAL ILLUMINATION

by A. A. KRUITHOF. 621.327.3 : 535 : 37

In connection with a general description of the principles and properties of tubular luminescence lamps printed earlier in this periodical, such lamps are here discussed which have been especially designed for use on the 220 volt mains. The most satisfactory dimensions of the lamps are derived from the maximum working voltage given, the desired light flux and an assumed limit for the brightness. Furthermore, after a brief description of the series apparatus the question of the colour of the lamps is carefully studied. Because of the great variety of tints which can be obtained by mixing different luminescent substances, it is important to know how the opportunities presented can best be exploited. The factors which here play a part are explained and illustrated with examples.

A light source of high efficiency can be obtained from the low-pressure mercury discharge which itself emits only little visible light and much ultraviolet radiation, when the ultraviolet radiation is converted into visible light by means of luminescent substances. The tubular luminescence lamps which have been constructed on this principle possess, in addition to their high efficiency, various other good qualities as well, which have already been discussed in this periodical [1]). The colour of the lamps exhibits practically no flicker, the light flux and the life of the lamps is much less dependent on mains voltage fluctuations than with the ordinary filament lamp. These properties give reason to expect that luminescence lamps will be widely used in the future.

In the article mentioned [1]) only one use of the lamp was described, namely for shops, restaurants, etc. where a large number of lamps in series are connected with a high-voltage transformer. This in many respects very simple and practical solution is naturally out of the question when luminescence lamps are to be used for more general purposes of illumination. In the home, for example, one or only a few lamps will generally be used at once, while moreover the use of high-voltage is impossible in many cases. Recourse must be had to the connection of each lamp with a separate series apparatus to the mains. This has certain consequences for the construction of the lamp which will be further considered in the following. At the same time the colour of the luminescence lamps will be discussed.

The construction of the lamp

The general construction of luminescence lamps has already been described in the article referred to [1]): The discharge takes place in a long glass tube on the inside wall of which the luminescent substances are deposited in powder form. The electrodes are in the shape of small spirals covered with a substance which emits electrons easily and which are kept at such a temperature by the bombardment by ions and electrons occurring in the discharge, that they are able to provide the electrons necessary to maintain the discharge. In the application previously described the length of the tubular lamps was 2 m (HTL 200), in the model for low voltage (TL 100), however, it has been reduced to 1 m with a tube diameter of 3.5 cm. This choice was made on the basis of the following considerations.

In order to make the lamp burn steadily at a mains voltage of 220 volts, it is found that the working voltage V_B may not be higher than about 120 volts [2]) —due partly to mains voltage fluctuations. The working voltage, aside from the voltage drop at the electrodes, is given roughly by the prod-

[1]) P. Schouwstra and G. Zecher, Tubular luminescence Lamps, Philips techn. Rev. 4, 337, 1939.

[2]) See: E. G. Dorgelo. Alternating-current circuits for discharge lamps, Philips techn. Rev. 2, 103, 1937.

uct of the gradient G (voltage per cm of the discharge-column) and the length l of the lamp. Now the gradient is mainly determined by the diameter d of the tube [3]), and the relation between G and d for a tube with the customary gas filling (mercury and argon) is given in *fig. 1.* If for V_B we assume

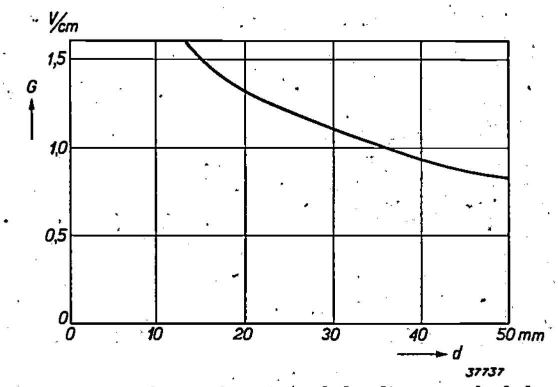


Fig. 1. Gradient G as a function of the diameter d of the discharge tube for the customary gas filling of the luminescence lamps (mercury + argon).

a certain value, then from fig. 1 for each value of d the required length $l = V_B/G$ of the lamp follows. This length is plotted as a function of d in *fig. 2*, for different values of the parameter V_B. The point l, d for the lamp to be constructed must in every case lie b e l o w the line $V_B = 120$ volts.

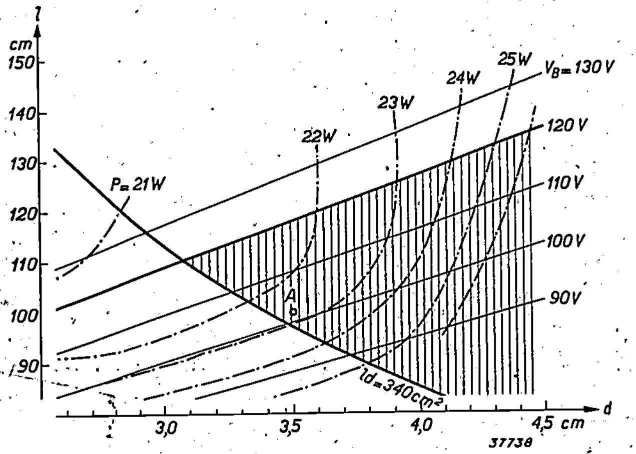


Fig. 2. The practically straight lines give the relation derived from fig. 1 between length l and diameter d of the tubular lamps for certain values of the working voltage V_B. The conditions $V_B<120$ volts and $B < 0.3$ c.p./cm² with a light flux $\Phi = 100$ decalumens limit the choice of dimensions for the tubes to the shaded area of the surface. The broken lines connect points at which for $\Phi = 100$ decalumens the same power P is needed. Point A gives the dimensions actually chosen.

A further limitation occurs due to the fact that on the one hand the lamp must give a certain minimum light flux — in our case this was fixed at $\Phi = 100$ decalumens —, while on the other hand it was considered desirable to choose such a low brightness that when the lamps were used without fittings no glare would be experienced from the direct light. With the high level of illumination to which one has become accustomed with modern lighting, even at the low height at which lamps are usually hung in private homes, it was found to be sufficient if the brightness B of the lamps was less than 0.3 cp/cm². For a tubular light source $\Phi = Bld\pi^2/10$. With $I = 100$ and $B<0.3$ it follows that $ld>340$ cm². If in fig. 2 we draw the line $ld = 340$, the point l, d to be chosen must lie to the right of this line.

The choice of the point l,d within the remaining shaded area of the surface is finally fixed by the desire to obtain as high an efficiency as possible. With a given light flux $\Phi = 100$ decalumens a certain light flux per metre tube length Φ/l corresponds to every point in fig. 2, and to this in turn a certain current i corresponds, which can be read off from the experimentally found group of curves in *fig. 3.* The variation of current and voltage of the

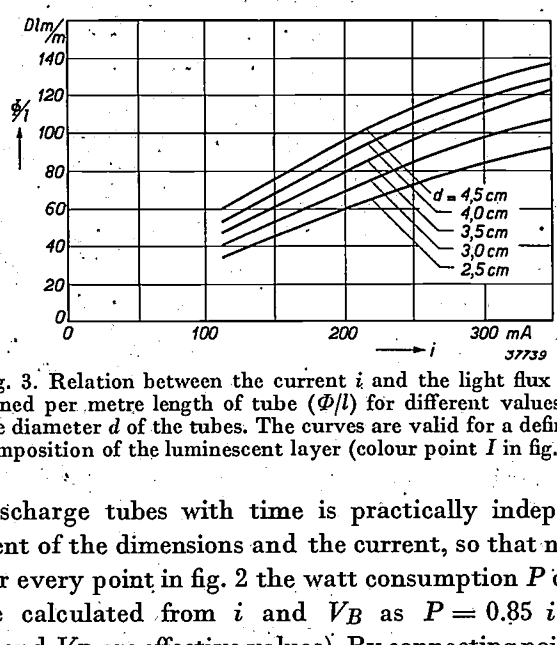


Fig. 3. Relation between the current i and the light flux obtained per metre length of tube (Φ/l) for different values of the diameter d of the tubes. The curves are valid for a definite composition of the luminescent layer (colour point I in fig. 9).

discharge tubes with time is practically independent of the dimensions and the current, so that now for every point in fig. 2 the watt consumption P can be calculated from i and V_B as $P = 0.85\ iV_B$ (i and V_B are effective values). By connecting points with equal value of P, the curves indicated by broken lines in fig. 2 are obtained. From the shape of these it follows that the point l,d can best be chosen in the extreme left-hand corner of the shaded part. The most economical lamp which satisfies our requirements as to V_B, Φ and B would thus have a length of 110 cm and a diameter of 3.1 cm. The length has been altered to 1 m, to which a diameter of 3.5 cm corresponds. The current for this point, indicated in fig. 2 by A is 250 mA, the wattage consumption about 22.5 W, the light yield about 44 lm/W.

[3]). To a certain degree the gradient still depends upon the current of the discharge, this influence is, however, neglected here for the sake of simplicity.

The series of curves of fig. 3 holds for a given composition of the luminescent layer in the tube. For other compositions of practical importance (other colours) the shape of the curves, and thus also that of the broken line curves in fig. 2, will be slightly different; the conclusion that the highest efficiency is obtained in the corner of the shaded area, however, remains valid. The most favourable dimensions for the lamp are therefore independent of the colour chosen.

The series apparatus

The series apparatus which is necessary to make the lamp burn steadily because of the negative characteristic of the gas discharge, consists in its simplest form of a choking coil. If we assume for this coil a reasonable consumption of about 5 W, the power factor at the above-mentioned values of the output and current in the lamp (22.5 W and 0.25 amp., respectively) is given by $\cos\varphi = (22.5 + 5)/0.25 \cdot 220 = 0.5$. In general therefore an improvement of the power factor will be desired in this case.

A point which requires special attention is the ignition. In the application previously described, where a high-voltage transformer was used, the no-load voltage of the transformer could easily be made so high that the tubes ignited under any circumstances. The mains voltage of 220 volts is, however, not immediately sufficient to cause breakdown in the gas of the tube. The ignition voltage must therefore be lowered by heating the hot cathodes of the lamp in advance, while in addition a voltage impulse is necessary. For this purpose the lamp is connected as in *fig. 4*. After the lamp is switched on with the main switch *1*, the auxiliary switch *2* is closed and a current of about 0.5 amp flows through the choke L and the two hot cathodes. These cathodes are of such dimensions that with this current they reach a sufficiently high temperature in about 1 sec. If the auxiliary switch is now opened, the self-induction causes a voltage impulse on the lamp whose magnitude depends not only upon the construction of the switch, etc. but also upon the phase of the alternating current at the moment of interruption. If this phase is by chance too unfavourable, the closing and opening of the auxiliary switch must be repeated.

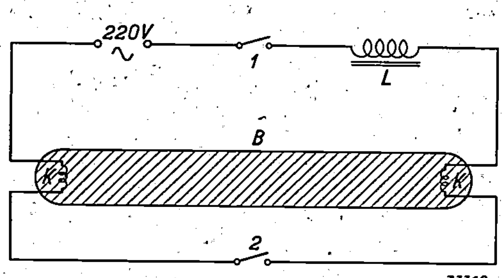


Fig. 4. Simplest connections of the luminescence lamps. B lamp, L choke, K hot cathodes, *1* main switch, *2* auxiliary switch for ignition.

The closing and opening of the auxiliary switch can take place automatically by constructing it in the form of a bimetallic switch of the type shown in *fig. 5*. The bimetallic strip B with the switch contact C forms one electrode of a small gas-discharge tube. When the mains voltage acts upon this due to the closing of the main switch (*l* in fig. 4) breakdown occurs in the tube, the bimetallic strip is heated by the discharge and, by bending, closes the contact C. The hot cathodes of the lamp are now heated. At the same time, however, the discharge in the small tube is short circuited, the bimetallic strip cools and the contact C is opened. If the lamp does not ignite the process is repeated until it does.

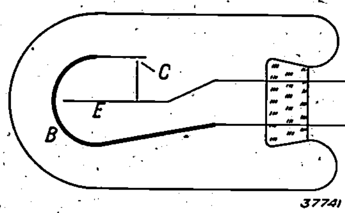


Fig. 5. Automatically working auxiliary switch for the connections in fig. 4. B bimetallic strip, E counter electrode, C switch contact. Between B and E a glow discharge occurs which heats the bimetallic strip.

The bimetallic switch may not of course break down when the lamp is ignited. The breakdown voltage of the switch tube must therefore be higher than the working voltage of the lamp and lower than the mains voltage. This can be realized by a suitable choice for the gas pressure in the small tube. The contact C and the other components must also be so constructed that the opening takes place abruptly enough to cause the desired voltage impulse.

In *fig. 6* a photograph is shown of the tubular lamp with a series apparatus consisting of choke and bimetallic switch.

Another method of connecting the lamp with the mains is with the use of a resonance connection such as is represented in *fig. 7*. If the self-induction L and the condenser C are in resonance at the mains frequency, then in every period a sufficiently high voltage acts on the lamp to cause it to ignite, especially when by means of suitable connections provision is made that upon switching on the hot cathodes are also heated. After ignition of the lamp it acts as a shunt on C, so that the resonance is disturbed. An advantage of the resonance connec-

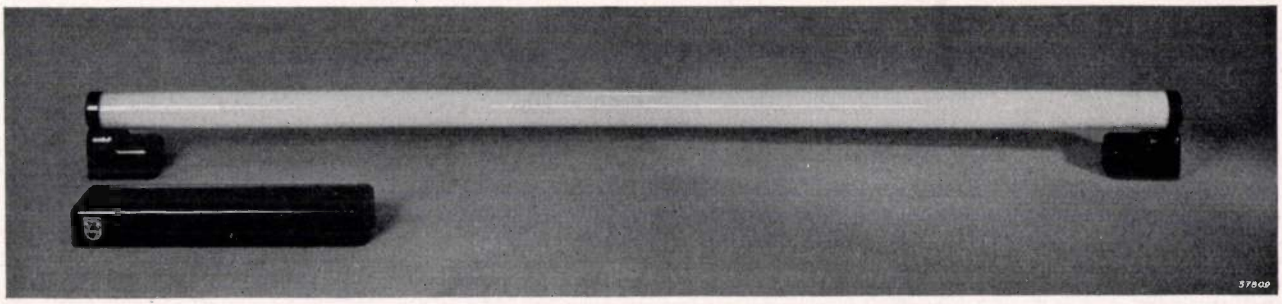


Fig. 6. Luminescence lamp TL 100 with series apparatus.

tions is that the lamp ignites almost immediately after being switched on, while, moreover, the connections can be so arranged that the power factor of the apparatus with the lamp burning is practically unity.

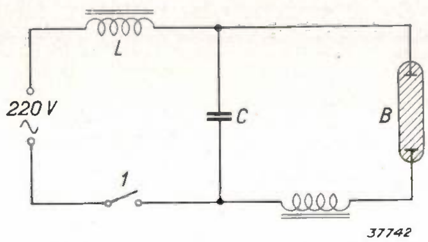


Fig. 7. Example of a resonance connection. The circuit formed by the choke L and the condenser C is tuned to the mains frequency.

The colour of the luminescence lamps

General considerations

One of the most important advantages of the luminescence lamp is the large degree of freedom it offers in the choice of the spectral composition of the light emitted, because of the possibility of mixing different luminescent powders (phosphors) in any desired proportion. In *fig. 8* the spectral energy distribution is given of the light of three of the most commonly used phosphors, namely of cadmium borate (red), willemite (green) and magnesium tungstate (blue). These colours are represented in the colour triangle of *fig. 9* by the points

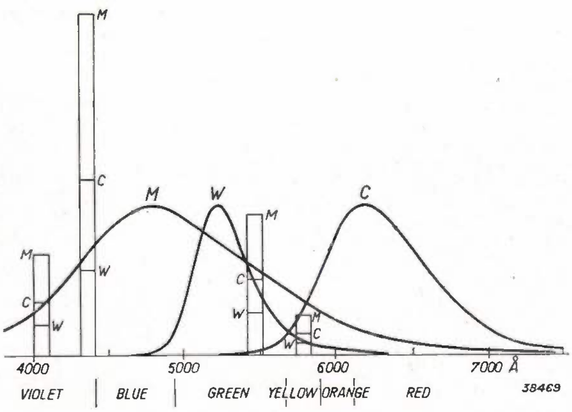


Fig. 8. Spectral energy distribution of the luminescence of cadmium borate (C), willemite (W) and magnesium tungstate (M). With the thickness of layer which gives the highest efficiency the visible mercury radiation simultaneously emitted by the lamps amounts to 10, 5 and 10 per cent, respectively, of the total radiation. The mercury lines are drawn exaggeratedly wide and of such a height that the area of the surface is a measure of the relative energy contribution at the thickness of layer mentioned. The ordinates for C, W and M stand in the correct relation for the case where the same total light flux is always produced. (The same holds for figs. 11*a* and 12*a*).

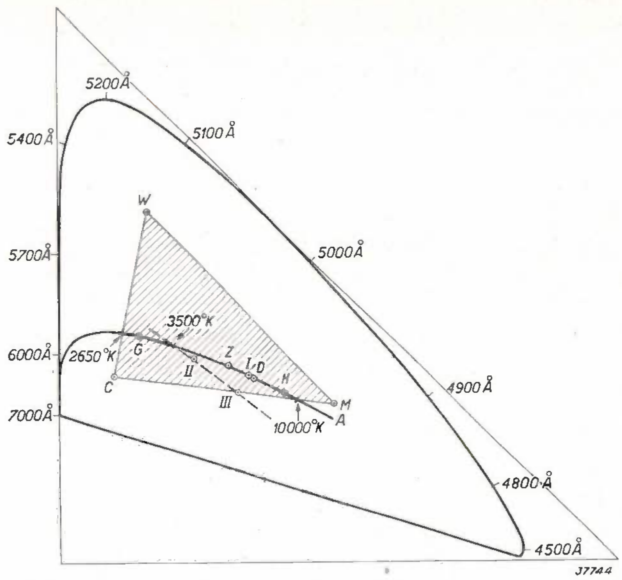


Fig. 9. Colour triangle. C, W, M are the colour points of the three phosphors calculated from the curves of fig. 8. By mixing the phosphors any colour point within the shaded triangle can be obtained, with an efficiency between 70 lm/W (at point W) and 35 lm/W (at C and M). On line A lie the colour points of black bodies at different temperatures (colour temperature). In addition the colour points are indicated of: gas-filled electric lamp G, sunlight Z, average daylight D, blue light of the sky H, and three kinds of luminescence lamps, I, II, III.

C, W, M. By mixing the three substances all the colours within the shaded triangle can be realized with an efficiency in the most favourable case (at point W) of about 70 lm/W, in the least favourable case (at point C or M) of about 35 lm/W [4]).

[4]) In the determination of the colour points for the three phosphors the fact must be taken into account that it is not exclusively luminescence light which is radiated by the tubular lamp, but that there is always a certain small proportion of mercury light. This proportion depends upon the thickness of the luminescent layer, which at the same

For the sake of comparison in fig. 9 the line A is also drawn upon which lie the colour points for glowing black bodies at different temperatures. Sunlight lies on this line (at point Z) with a colour temperature of about 5 000 °K (*i.e.* its colour corresponds to that of the radiation of a black body at about 5 000 °K), and furthermore so-called average daylight (the light when the sky is completely overcast), (point D) with a colour temperature of about 6 000 °K, the light of the blue sky H with a colour temperature of about 8 000 °K and an ordinary gas-filled electric lamp of for instance 220 volts, 500 W, point G, with a colour temperature of 2 850 °K. By using filaments at different temperatures other points on line A could also be realized with the electric lamp; at points to the right of G, however, their life would be short and to the left of G their efficiency, which at G itself may be about 16.5 lm/W, decreases rapidly. This is also the case when one attempts to reach points to the right of G with a filament at normal temperature combined with filters, for instance in the form of a coloured bulb; the Philips "sunlight lamp" which is constructed on this principle and whose colour temperature is about 4 200 °K, has for instance only half the efficiency of a normal gas-filled lamp.

How can we make the best use of the great freedom in the choice of colour of the luminescence lamps?

Considering the fact that according to fig. 9 all colour temperatures between 2 500 and 10 000 °K can be obtained by a suitable mixing of the phosphors, an obvious solution would be to use the luminescence light with its excellent efficiency to replace or supplement daylight or electric light. In doing this, however, various factors must be taken into account.

In the first place at a given level of illumination it is found that the colour temperature must lie within certain limits if the effect of the illumination is to be pleasing. Roughly, it may be said that a low or a high colour temperature corresponds to a low or a high level of illumination, respectively. We have investigated this relation experimentally somewhat more closely by introducing in a room a variable number of electric lamps whose current (*i.e.* the temperature of the filaments) could be varied. The result is given in *fig. 10*, while in the text below the figure the experiments are described. Below the lowest curve the illumination is "dim" (at low colour temperature) or "cold" (at high colour temperature). Above the highest curve the unnatural colour reproduction was unpleasant. These obviously vague limits within which the illumination is considered "pleasing" could in our experiments be determined at least with an accuracy of 20 or 30 per cent.

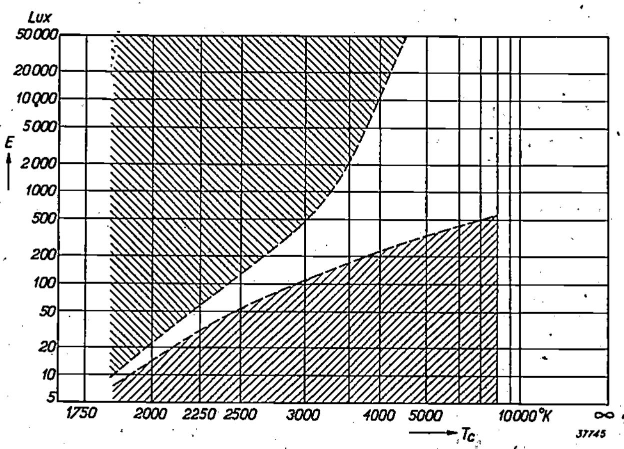


Fig. 10. For every colour temperature there exists a highest and a lowest level of illumination at which the illumination is considered "pleasing": at lower levels the illumination appears dim or cold, at higher levels the colour rendering is unnatural. The left-hand part of the limiting curves, up to a colour temperature of 2 850 °K, is recorded by allowing electric lamps with variable (decreased) current to burn in a room, and varying the number of lamps. The illumination intensity on a table 80 cm high was here measured. In the right-hand part the lowest level which does not give the impression of coldness was determined by experiments with daylight itself and with the daylight luminescence lamps to be described below. The shape of the upper curve has been extrapolated in this region with the help of the fact that in direct sunlight (colour temperature 5 000 °K) even with the highest illumination intensities occurring (10^4 or 10^5 lux) the colour rendering is never found "unnatural". On the abscissa the reciprocal value of the colour temperature T_c is plotted, on the ordinate the logarithm of the illumination intensity E, since $1/T_c$ and $\log E$ are measures of the physiological estimation of these quantities. In these coordinates the lower limiting curve takes on a nearly linear form. It may be mentioned that the experiments were carried out in a laboratory room. It was found, however, that in a living room with light-coloured furniture and wall coverings roughly the same limits are obtained.

From fig. 10 various conclusions may be drawn. If for interior illumination an approach to daylight is desired by the use of luminescence lamps with high colour temperatures, then according to the curve at a colour temperature of 4 000 °K an illumination intensity of at least 240 lux must be used (see the text under fig. 10 for the significance of these figures), at 7 000 °K at least 500 lux. Since the level of illumination is already determined by other considerations the colour temperature cannot be chosen arbitrarily high; if in a living room, for example, one does not wish to use more than 150

time determines the efficiency. In each case a thickness of the layer is assumed such that the highest possible efficiency is obtained; the percentages of mercury light hereby occurring for the three phosphors are indicated in fig. 8, and from the spectral composition of the total radiation the three colour points are then calculated. On the general subject of the colour triangle see for example Philips techn. Rev. **1**, 282, 1936 and **2**, 39, 1937.

lux, then according to the figure 3 500 °K is the upper limit for the colour temperature.

A further limitation occurs in the special case where the new light source must be capable of being combined with that already existing, namely with daylight or electric light. The kinds of light to be combined will then have to have approximately the same colour (the same colour temperature), or at least they must exhibit no unpleasant contrast of colour. But in addition the colour rendering obtained with the different light sources must also not be too different. Similarity of colour by no means ensures similarity of colour rendering, since the latter is determined by the spectral composition of the light and the same colour point in the colour triangle can be obtained with quite different spectral compositions. The colour rendering will be dealt with in more detail below.

While until now we have only considered cases in which the luminescence lamps were to be used for imitating or in combination with radiations on the line A in fig. 9, the question may now be put as to whether luminescence lamps may not also be considered which have colour points in other parts of the shaded triangle. The question whether such a choice of the colour point can have special advantages we shall leave open [5]).

It is indeed found that a fairly large deviation from line A is permissible if the following points are borne in mind. In the first place there is the seemingly paradoxical requirement that the light must not appear definitely coloured. Upon direct comparison of course all kinds of light appear in general coloured. If, however, there is no possibility of comparison, daylight and electric light do not appear coloured but "relatively white", and this is also true for kinds of light with colour points not on line A as long as the deviations are not too great.

The second requirement concerns the colour rendering. While the colour rendering can be judged by comparison when luminescence is used in combination with other light and the designation of the colour impression obtained and the saturation of the colour obtained must agree, when only luminescence lamps are used no comparison is possible. In judging the colour rendering therefore in this case one must have recourse to "colour memory" which is chiefly confined only to the designation of colours. Care must therefore be taken that the classification of the shades of colour of the objects in our environment in the light of the luminescence lamps does not become unnatural, *i.e.* does not deviate too much from that in daylight.

In the third place there is now also a limitation of the colour temperature in connection with the intensity of illumination used. At colour points which do not lie on the line of black bodies we may not indeed speak of a colour temperature. Nevertheless as experiments by Judd [6]) have shown we can "assign" colour temperatures to such points without ambiguity; this temperature is then that at which a black body exhibits the smallest possible colour difference with the kind of light under consideration. Experiments have shown that fig. 10, at least as far as the lower limiting line is concerned, also gives useful results for assigned colour temperatures.

Examples

We shall now illustrate the way in which the results obtained can be put into practical use.

As first example we shall consider a lamp which can be used specially for supplementation or taking the place of daylight. Since the colour temperature of daylight varies between about 5 000 and 8 000 °K, according as one is concerned with direct sunlight or with the light of a more or less overcast sky, the colour temperature of the lamp must be chosen between about 5 500 and 7 000 °K if it is not to exhibit a large colour difference with sunlight. In order not to be compelled to use too high illumination intensities when the lamp is used alone, the colour temperature should not be chosen higher than 5 500 °K. The colour point of this lamp then lies at I in fig. 9. The spectral energy distribution of the light which is obtained with the mixture of the three phosphors corresponding to this is shown in *fig. 11a*. In *table I* the composition of the light is indicated according to the block method [7]) often used in this periodical; the whole spectrum is hereby divided into eight wave length regions (blocks), whose proportions of the total light radiation are measured or calculated. A comparison with the figures for average daylight given in the same table shows that a good agreement in the colour rendering with the two kinds of light can also be expected. This was tested experimentally with the help of the colour cards of Ostwald's colour atlas, in three steps. First the brightness of the cards in

[5]) We confine ourselves here expressly to normal interior illumination. It is unnecessary to state that special effects, for instance for advertising purposes, can be obtained by deviation from line A.

[6]) D. B. Judd, J. Opt. Soc. Amer. **26**, 421, 1936.

[7]) See for example Philips techn. Rev. **4**, 66, 1939.

Table I

Distribution of the light of different light sources over the eight blocks of the spectrum indicated

	1 (4 000–4 200)	2 (4 200–4 400)	3 (4 400–4 600)	4 (4 600–5 100)	5 (5 100–5 600)	6 (5 600–6 100)	7 (6 100–6 600)	8 (6 600–7 200)
Lamp *I*	0.016	0.432	0.638	9.3	42.9	36.8	9.43	0.363
Daylight	0.027	0.228	0.801	10.8	40.8	36.3	10.3	0.73
Lamp *II*	0.010	0.409	0.224	4.06	38.4	43.4	13.0	0.471
Electric lamp	0.005	0.060	0.249	5.47	33.4	42.6	16.7	1.57
Lamp *III*	0.012	0.475	0.462	7.25	35.0	38.1	18.0	0.695

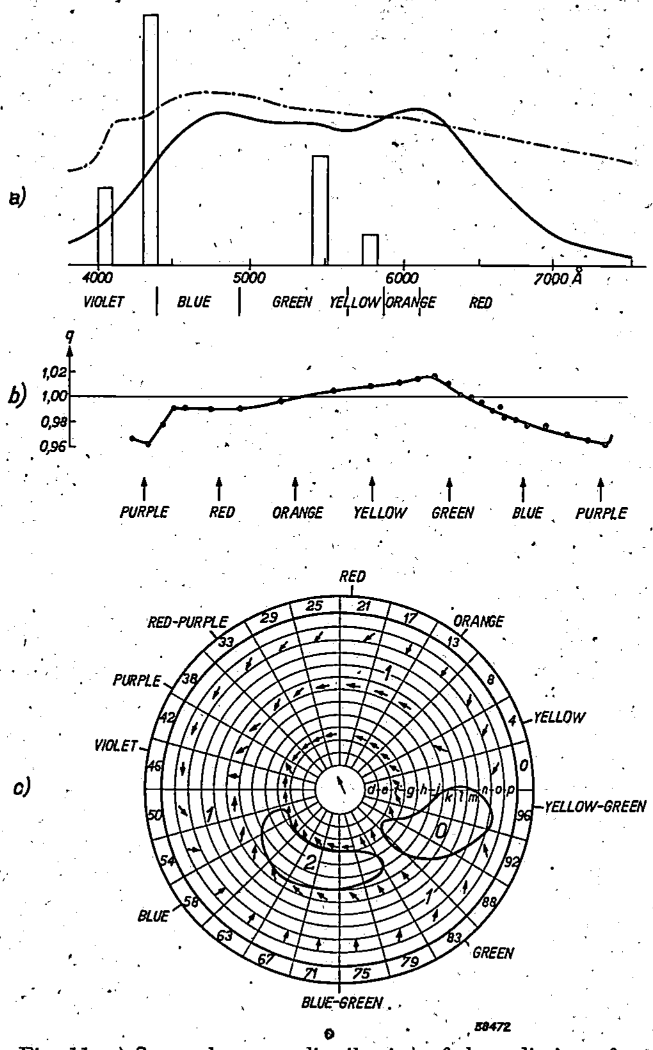


Fig. 11. *a*) Spectral energy distribution of the radiation of a luminescence lamp in which the three phosphors *C*, *W*, *M* are so mixed that the colour point *I* in fig. 9 is obtained. The broken line curve is for average daylight.

b) A coloured card upon illumination with average daylight in general exhibits a different brightness than upon equally intense illumination with the luminescence lamp *I*. For 24 colour cards of the *nc* series of Ostwald's colour atlas, of the colours given along the abscissa, the ratio *q* between the two brightnesses measured is plotted. At all points *q* lies so close to 1 that the differences are practically unnoticeable.

c) A complete set of colour cards (*c*-set) which was first observed in daylight and then in the light of lamp *I*. Each compartment represents one card. The colours are designated around the circumference of the circle (numbered 0-96); from the centre toward the outside the saturation increases (letters *d* to *p*). The arrows indicate the colour shifts observed. Only in the region marked "2" are such shifts clearly observable.

the two kinds of light to be compared was determined from the spectral variation of the coefficient of reflection of the cards and from the spectral distribution of intensity of the two kinds of light. For a number of cards (of the *nc* series) the ratio of the two brightnesses calculated is plotted in fig. 11*b*. It may be seen that this ratio does not deviate very much from unity for any of the cards. As second step the shade of colour and saturation were investigated for a series of cards by illuminating one half of each card with the luminescence light and the other half with average daylight [8]). The difference in colour between the two halves was designated by 0 (not appreciable), 1 (just appreciable), 2 (clearly appreciable), 3 (very striking, quite different colour). From the results given in fig. 11*c* it may be seen that the differences are almost everywhere limited to "1"; there is only a very small region with differences of the order of "2"; while differences "3" do not occur. Most colours are somewhat less saturated in luminescence light than in daylight. Finally as third step, experiments were carried out in which the colour rendering was judged exclusively by colour memory: the colour cards of the series mentioned were illuminated with the lamp and an attempt was made to name the colours observed as accurately as possible. The differences occurring with respect to the naming of the colours observed in daylight were found to be slight and in satisfactory agreement with the diagram in fig. 11*c*.

Summarizing it may therefore be stated that the luminescence lamp with the colour point *I*, both in respect to its colour and in respect to its colour rendering, very closely approaches average daylight and may be used not only to replace but also to supplement the latter.

If the emphasis lies more in the supplementation than in the substitution, then it is not actually necessary to try to approach average daylight so

[8]) This method has already been described by P. J. Bouma, Colour reproduction in the use of different sources of "white" light, Philips techn. Rev. 2, 1, 1937.

closely. A supplementation of daylight will in practice be necessary chiefly during the last few hours before sunset, at which time the colour temperature of daylight, expecially with overcast sky generally lies lower [9]) than the average value (the light is redder). The colour temperature of the lamps may thus be chosen somewhat lower.

In the second example which we shall discuss it was mainly a question of low levels of illumination. For reading and other work an illumination intensity of 150 to 250 lux is often applied. According to fig. 10 the colour temperature must then lie between about 2 550 and 4 100 °K. The lower limit mentioned corresponds to a colour point in the colour triangle of fig. 9 which is obtained by mixing practically only cadmium borate and willemite (line *C-W*). A glance at the spectral distribution in fig. 8 shows that the light so obtained will contain practically no blue or violet except what is present in the mercury light (line 4 358 Å). It may therefore be expected that the colour rendering by a lamp with such a low colour temperature will not be satisfactory. It is thus better to choose for the purpose in view a colour temperature closer to the upper limit, 3 500 °K, for example, where the content of blue in the light of the magnesium tungstate improves the colour rendering. Further improvement can still be attained by deviating slightly from the line *A* for black bodies and adopting colour points with the assigned (see above) colour temperature of 3 500 °K. These points all lie on the dotted straight line of fig. 9, and the point to be chosen on this line is indicated by *II*. *Fig. 12a* shows the spectral energy distribution of the light of this lamp, while in table I the block division is given. For the sake of comparison the spectral distribution of intensity of the electric filament lamp is also given here. The colour rendering of lamp *II* has been investigated in the same way as described above, and primarily by colour memory. Only in the orange region were there are appreciable differences from daylight. The differences lie, however, in the same direction as those which occur upon illumination by ordinary electric light, to which we are already accustomed. A direct comparision of the colour rendering of lamp *II* with that of the electric lamp produced quite considerable differences, see fig. 12*b*. As was to be expected the colour rendering of lamp *II* is more nearly like that of daylight. In spite of this difference and the difference in the colour of the light itself from that of the ordinary electric lamp, the effect obtained upon combination of these two kinds of light is not found to be too unpleasant. As far as the difference in colour of the light is concerned this is because of the fact that the line joining the points *II* and *G* is practically parallel to line *A*; it is a well known phenomenon that such a difference in colour (blue-yellow contrast) is experienced as much less unpleasant than the difference in colour between two points whose connecting line is about perpendicular to line *A* (red-green contrast).

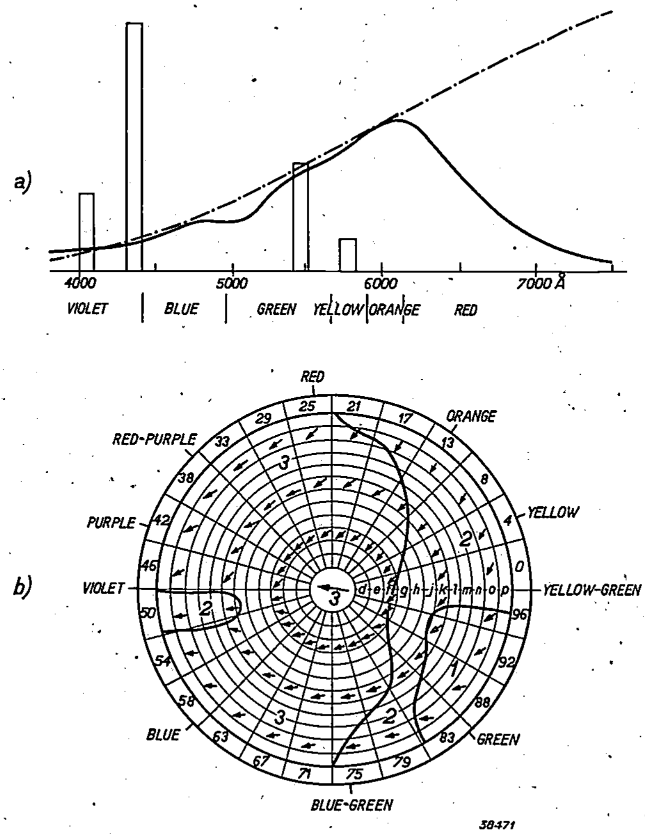


Fig. 12. *a*) Spectral energy distribution for luminescence lamp with colour point *II* in fig. 9. The broken line curve refers to ordinary electric light.
b) Colour shifts of the colour cards upon illumination first with electric light, then with lamp *II*.

Finally an example may be given in which use was made of the possibility discussed above of using colour points which are fairly far away from the line for black bodies. The advantage of such a light source may consist in a modification of the colour rendering in the sense that the colours of all objects appear more saturated, so that the whole interior under this illumination makes a fresher and brighter impression. A similar effect can be obtained for instance with the ordinary electric lamp by decreasing the energy contribution of the block 5 600—6 100 Å by means of a filter of

[9]) See S. S. Ornstein, J. G. Eymers, D. Vermeulen and G. W. Postma, Daylight measurements in Utrecht, Proc. Kon. Akad. Wet. Amsterdam 16, 1—79, 1936.

"Philiphane" glass[10]). By daylight a similar effect is produced when the energy contributions of the blocks 5 100—5 600 and 5 600—6 100 Å are both decreased. With a certain mixture of phosphors (colour point *III* in fig. 9) a spectral energy distribution with such a minimum in the wave length region in question can be obtained, see *fig. 13.* The lamps constructed with this mixture also actually produced for many colours the effect of the increase in saturation. The condition that the lamp itself must still appear "relatively white" is here no longer satisfied. The impression of being coloured is, however, still very weak, and in connection with the special applications of this lamp it is not serious. The colour rendering must in this case be judged exclusively by colour memory, since the lamp is not intended for use in combination with other sources of light. The designations of the colours found deviate here also only little from those by daylight.

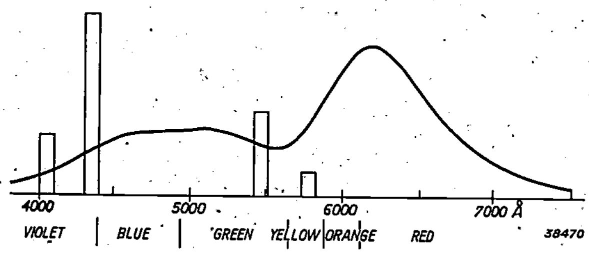
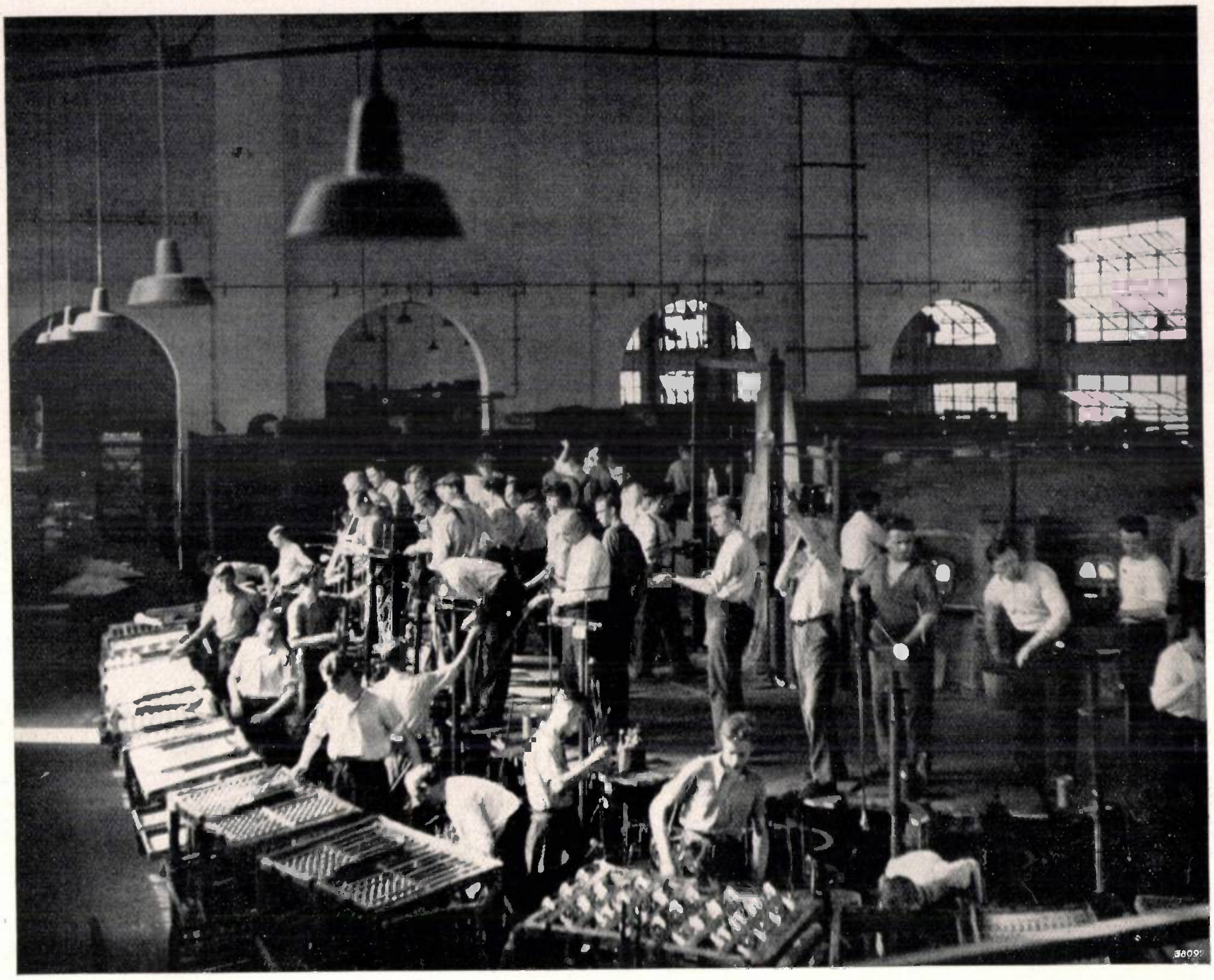


Fig. 13. Spectral energy distribution for the luminescence lamp with the colour point *III* in fig. 9.

[10]) P. J. Bouma, The colour reproduction of electric lamps and "Philiphane" glass, Philips techn. Rev. **3**, 46, 1938. It may be pointed out that in this discussion the term "saturation" must not be taken in a physical but in a physiological sense; a colour gives the impression of being more saturated according as the colour point is farther away from the colour point of the light used.

VIEW OF ONE OF THE HALLS OF THE PHILIPS GLASS FACTORY

At the beginning of this year the glass factory had been in existence for 25 years.



While the manufacture of the bulbs for ordinary electric lamps and radio valves is at present mechanical, the blowing of bulbs for the many kinds of special lamps is still one of the branches of the industry where the old handicraft has been retained. The glass blowers stand in a ring around the glass furnace in which are situated the pots of molten glass. They take out portions of glass on the long pipes and blow it into the desired shape. The pipe with the still hot bulb is then handed over to the boys who form a second ring around the glass blowers, and who knock off the bulbs and place them in the racks. The picturesque impression which is given by the whole does not detract from the great efficiency and precision with which the work is actually done.

A SIMPLE APPARATUS FOR COUNTING ELECTRONS

by A. BOUWERS and F. A. HEIJN.

539.16.08

For the measurement of electron beams or of γ-rays an electron counter is usually used which contains as its most important component a gas-discharge tube fed with a D.C. voltage of about 1 000 volts which gives a current impulse for every electron entering (Geiger counter). After a brief explanation of the general construction of counting tubes and the connections in which they are used, a simple counter is described which is especially intended for use in demonstrations, for orienting measurements and the like. In contrast to precision counters which are manufactured by Philips for scientific purposes, in the apparatus described here emphasis has been laid upon obtaining a unit which is simple in operation and easy to transport. This has been made possible partly by the application of a cascade connection consisting of small condensers and selenium valves for exciting the necessary D.C. voltage.

The processes which take place in the nuclei of atoms are often manifested by some kind of radiation which accompanies them. The classic example is the spontaneous decomposition of radium in which not only electromagnetic waves (γ-rays) are emitted, but also material particles (electrons or β-rays and helium nuclei or α-rays). The experimental investigation of such processes indeed amounts chiefly to the detection and measurement of the radiations emitted.

The methods developed for this purpose have already reached a high degree of perfection. For example, with the help of the Wilson chamber and a magnetic field it is possible to determine the velocity of emitted electrons with great accuracy. With ionization chambers and amplifiers extremely accurate measurements have been carried out on the range of heavy particles (hydrogen nuclei for example).

In many cases, however, it is not a question of carrying out such precision measurements. For demonstrations during lectures, etc. and for orientating tests an instrument will suffice with which the presence of the radiation alone can be shown, and its intensity estimated if desired, while in these cases it is of more importance to make the apparatus as simple as possible and preferably easy to transport.

In this article an instrument will be described for the detection — the "counting" — of electrons, which has been especially developed for the use mentioned. For a better understanding it is desirable that something should first be said about electron counters in general.

Principle of the electron counter

For detecting an electronic radiation use is ordinarily made of the Geiger counter. This is a gas-discharge tube with a certain electrode configuration which can for instance be connected as indicated in *fig. 1*. A D.C. voltage, slightly higher than the normal breakdown voltage is applied to the tube. Breakdown only occurs, however, when in the gas of the tube there is a "germ" in the form of a free electron. On its way to the anode this electron will ionize a number of atoms; each of the electrons freed in this way repeats the process and an avalanche of electrons ensues which initiates an independent discharge. With a suitable shape of tube and a sufficiently large resistance R in fig. 1, however, the discharge is found to be unstable: it breaks off after a very short time, after which the entire phenomenon can be repeated with a new "germ". The germs may now be provided by an electronic radiation which enters the discharge tube (counting tube) from the outside. An electron then manifests itself by a short current impulse through the tube which may be observed and "counted" with an electrometer as indicated in fig. 1.

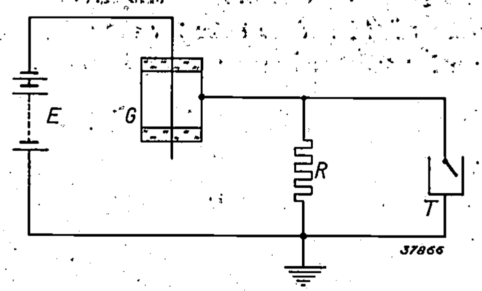


Fig. 1. Simple connections for the Geiger counter. G counting tube, R resistance of at least 10^9 ohms, E D.C. voltage source for 1 200 volts, T electrometer or other instrument for recording the current impulses.

A suitable electrode configuration for the counting tube has been found to be a cylindrical cathode along whose axis a fine wire is stretched as anode. The gas pressure is made several cm Hg, the voltage 900 to 1 200 volts. For the rest, different investigators have given very different (sometimes contradictory) specifications about the nature of the gas to be chosen, the materials of the electrodes, etc. This lack of agreement is a reflection of the fact that no explanation satisfactory in every respect has yet been found for the mechanism of the phenomena occurring in the counting tube, in particular for the breaking off of the discharge [1]). For this breaking off it seems to be essential that the current which begins to flow after the breakdown in the tube should not exceed a certain very low value. The current which increases rapidly upon breakdown causes a rapidly increasing voltage drop along the resistance R, so that the voltage on the tube falls. If R is larger, then already at a very small current the voltage loss becomes several hundred volts, with the result that the current does not exceed the above-mentioned critical value and the discharge is extinguished.

In the simple connections sketched in fig. 1 R must be at least of the order of magnitude of 10^9 ohms to obtain the desired effect. This results in very high requirements, which are difficult to satisfy in practice, being made of the insulation of the electrodes and connections; the insulation resistance must of course be greater than R, since it is in parallel with it. Another disadvantage of the high value of R is the following. If two electrons enter the counting tube in quick succession, the second electron will only cause a separate current impulse when the counter has already "forgotten" the impulse due to the first electron, *i.e.* when it has returned to its initial state. This takes a certain time, and this time is largely determined by the speed with which the voltage on the tube rises to the original value after the breaking off of the discharge. The RC time of the current circuit provided a measure of this. If we take for the capacity C of the counting tube and connections a value of 10 $\mu\mu$F — in practice one can scarcely reach a much lower value —, then at $R = 10^9$ ohms the RC time is 10^{-2} sec. This is therefore approximately the limit of the "resolving power" of the counter, *i.e.* two electrons which enter the counting tube within a time interval of less than 10^{-2} sec are counted as a single electron. The resolving power is obviously lower, the larger the value of R.

Considerable improvement with respect to the required magnitude of R is obtained by means of the connections indicated in *fig. 2* [2]). The voltage on R is applied to the grid of an amplifier valve which in the normal state passes no anode current due to a sufficiently strong negative grid bias. With a very small voltage on R, caused by the ignition of the counting tube, a large anode current begins to flow, which causes such a voltage drop along the anode resistance R_0 that the tube is extinguished. In order to bring about the extinction in the same way as in the connection of fig. 1 (*i.e.* at the same value of the current through the tube) the resistance R need now be only about 10^7 ohms, with $R_0 \approx 10^6$ ohms. Then a normal insulation of the electrodes and connections suffices, while the resolving power of the counter has become one hundred times as great.

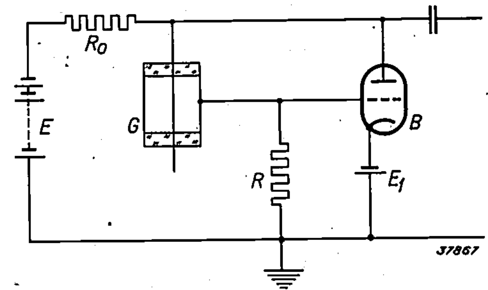


Fig. 2. Connections according to Neher and Harper. The voltage on R is amplified by an amplifier valve B. The resistance R here needs only to be 10^7 ohms, the anode resistance R_0 much less still. The D.C. voltage source E_1 provides the necessary grid bias.

We must stop a moment at the last point. Besides the counting tube and the connections belonging to it, the apparatus also contains an indicating instrument, for which as the simplest possibility we have assumed an electrometer in the above. Instead of this an instrument is usually used which automatically records the number of current impulses received, namely the familiar call counter of telephone technique. It is clear that the resulting power of the electrometer or call counter is always fairly limited, due to their mechanical inertia. In general it is much less than 100 per second. And since it is true here also that the slowest ship determines the speed of the whole fleet, it would be useless to increase the resolving power of the counter connections as much as possible without at the same time arranging the indication on the same basis. This can be done with the help of an automatic counting circuit ("scale of two" [3])) which transmits only every second current impulse of the counting tube to the call counter or to a second, third, ... counting circuit which works in the same way. By carrying out the counting in four stages, for instance, only every 16th electron is recorded in the call counter, so that it can handle 16 times as many electrons per second as its mechanical inertia would permit.

[1]) A survey of a number of important publications on this subject is given by J. H. Gisolf: Elektronentellers, Ned. T. Natuurk. 4, 129, 1937.

[2]) H. V. Neher and W. W. Harper, Phys. Rev. 49, 940, 1936.

[3]) The first arrangement designed for this purpose was announced by C. E. Wynn-Williams, Proc. Roy. Soc. A **136**, 312, 1932.

Electron counters with such counting circuits, which are indispensable for precise investigations, are also manufactured by Philips. Of the apparatus here described, however, an especially high resolving power was not required. The connections of fig. 2 were therefore used which make all the special precautions as to insulation superfluous (keeping the air dry, etc.), no counting circuit is used, however, in order not to make the apparatus unnecessarily complicated and heavy. We shall now study the apparatus itself in somewhat more detail.

Description of the apparatus

The counting tube

The most important structural data of the counting tube have already been mentioned above, while it was also stated that in certain respects, especially with regard to the material of the electrodes and the gas filling, opinions as to the best choice are quite divergent. A series of investigations on these points have been carried out in this laboratory [4]), and finally led to the result that the choice of the electrode material and of the gas are not so important as was previously thought, if care is taken to secure the greatest possible purity. This means among other things that the anode wire and the walls of the counting tube may not contain any adsorbed impurities which could be liberated when the tube is in action. The counting tube must therefore be carefully heated and degassed before it is filled with the desired gas (a mixture of argon and nitrogen, for example) and sealed off, just as is done with transmitter valves, X-ray tubes, etc. In the case of the counting tubes here developed, just as in that of the Philips transmitter valves and X-ray tubes, this is made possible by the use of chrome-iron: by making the cylindrical cathode of this material it is possible to weld a glass tube to both ends of it to support a tungsten wire serving as anode. The welds are vacuum-tight and heat-resistant. The construction can clearly be seen in the X-ray photograph *fig. 3*. The cathode cylinder is made so thin (about 0.1 mm) that electrons with very low speeds can still pass through it. The 0.1 mm thick anode wire is heated during the degassing by the passage of a current. In order to keep the wire tightly stretched during this process and later also (it must lie along the axis of the cathode cylinder) the spring visible in fig. 3 on the left is introduced.

It is interesting to note that the glass used for the counting tube may contain no potassium, since this element exhibits a natural, although very weak, radioactivity, and the electrons emitted would continually give rise to current impulses. Another cause of these undesired current impulses is the cosmic radiation which is present everywhere in space. Since the number of impulses occurring due to this cause is proportional to the volume of the counting tube, the tube has been made fairly short for our apparatus. (Conversely, the counting tubes which are made by Philips especially for investigating cosmic rays are very large).

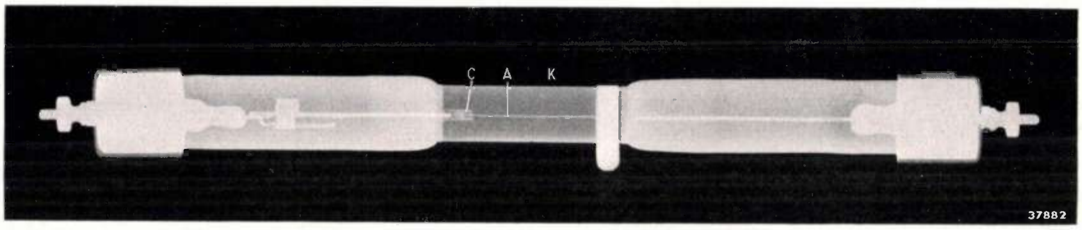


Fig. 3. X-ray photograph of the counting tube. At both ends of the thin-walled chrome-iron cathode cylinder K a tube of good insulating potassium-free glass is welded. The spring visible on the left keeps the thin tungsten wire A, serving as anode, tightly stretched. The shielding tubes C serve to improve the shape of the electric field in the tube.

The quality of a counting tube can be judged by means of the so-called count characteristic which indicates how many electrons the tube counts at different voltages when a given radiation, *i.e.* a given number of electrons per minute, is allowed to enter it. At very low voltage (too low for the occurrence of electron avalanches) the tube will obviously show no reaction; at very high voltage the fall in voltage caused by the resistance R will not be able to cause extinction, and a continuous discharge will therefore occur. Between these two extremes lies a more or less sharply-defined voltage region, the "counting region", within which the tube gives a number of current impulses proportional to the number of electrons entering it. The wider the counting region the better the counting tube. In *fig. 4* a characteristic is given of the counting

[4]) W. de Groot and J. H. Gisolf, Ned. T. Natuurk. **3**, 161, 1936; J. H. Gisolf, Physica **4**, 69, 1937, and the article referred to in footnote [1]).

tubes here described. It may be seen that, with the radiation used for recording, the voltage may lie between about 1 000 and 1 150 volts. With more intense radiation the counting region is smaller.

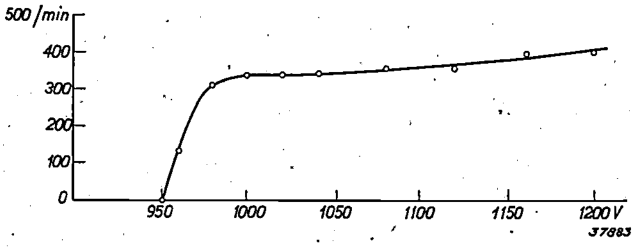


Fig. 4. Count characteristic of the counting tube. A sample of 1 mg of radium enclosed in a lead case 1 cm thick and placed at a distance of 1 m from the counter served as source of radiation. The number of current impulses recorded by the counter is plotted as a function of the voltage on the counting tube. The "counting region" within which the number of current impulses provides a correct measure of the intensity of the radiation extends in this case from about 1 000 to 1 150 volts.

The connections

The connections of the apparatus are given in *fig. 5*. They correspond for the most part to those of fig. 2. As amplifier valve B_1 a valve with a steep slope is used, which gives a high anode current and thus a large fall in voltage on the counting tube with a small increase in the grid voltage. The necessary grid bias cannot in this case, as is usual in ordinary amplifier connections, be obtained with a resistance in the cathode connections, since the valve carries no current in the no-load state. Therefore the bias is taken by means of a potentiometer from the plate voltage apparatus which serves for the supply. For the counting of the current impulses they are further amplified by a second amplifier valve B_2 and applied to a call counter, the rattling of which already makes it possible to estimate the intensity of the radiation by ear without reading it off.

As was stated above, a D.C. voltage of from 900 to 1 200 volts is needed for the counting tube. The generation of such a voltage is of itself no problem; according to the ordinary method, namely the use of a transformer with rectifier valves, fairly heavy constructions are, however, obtained. In order not to load the apparatus with these, a different method has been chosen here. In this periodical [5]) it has already been explained that with the help of a connection of condensers and valves (cascade connection) from an A.C. voltage with the peak value E a D.C. voltage $2E$, $3E$, $4E$, etc. can be obtained. This connection, which is in general only used for obtaining very high voltages (of hundreds of kilovolts), is found to be advantageous in this case also for the relatively low voltage of 1 200 volts. The A.C. voltage is provided by the mains (at 220 volts the peak value is about 300 volts), and can be further regulated with a potentiometer. The four successive stages of condensers and rectifiers with which the high D.C. voltage is obtained are shown in fig. 5. As rectifiers selenium valves are used which in order to obtain the required blocking voltage of 600 volts are in each case built up of about 20 elements piled one on top of the other, each with a blocking voltage of about 30 volts [6]). In this way an extremely small and light unit is obtained as shown in *fig. 6*.

In *fig. 7* a photograph is given of the complete apparatus which can be connected directly to the light mains.

[5]) Philips techn. Rev. 1, 6, 1936.

[6]) On the subject of selenium valves and their applications see D. J. Duinker, Philips techn. Rev. 5, 199, 1940. The selenium elements here used are of a special construction for very small currents.

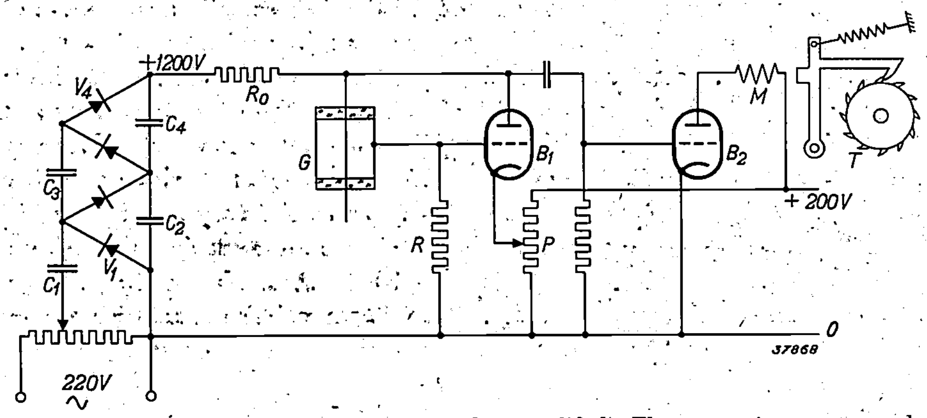


Fig. 5. Connections of the apparatus (somewhat simplified). The connections correspond for the most part to those of fig. 2. The current impulses caused by the counting tube G in the amplifier valve B_1 are amplified in a second valve B_2 and fed to a counting device (call counter) T. This contains a magnetic coil M excited by the anode current of B_2, which at every current impulse attracts a lever which causes a toothwheel to move by means of a hook. The supply voltage for the counting tube is obtained by means of a cascade connection of condensers C_1-C_4 and selenium valves V_1-V_4. The grid bias of the amplifier valve B_1 is regulated with the potentiometer P.

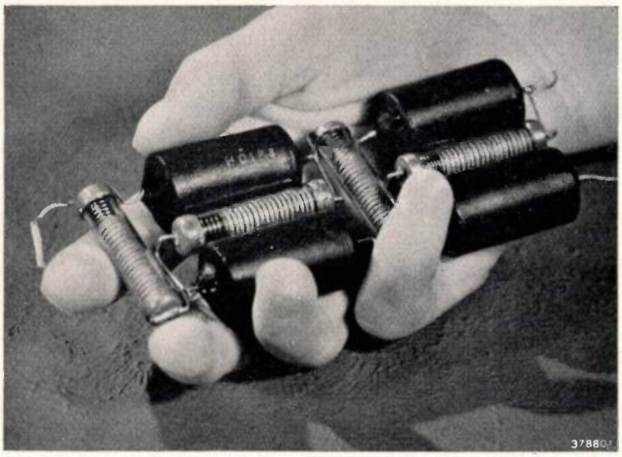


Fig. 6. Cascade connections for 1 200 volts. Each of the valves consists of 20 selenium elements piled one on top of the other in a glass tube and pressed firmly together by a spring.

In conclusion a few words must be said about the possibilities of application. Besides for the applications already mentioned, such as the detection of an electronic radiation for demonstration purposes, the apparatus can also be used in other fields. In hospitals it sometimes happens that radioactive substances which were enclosed in bandages for irradiating a patient are lost. They can often be found by going over the room in which they are thought to be present with an electron counter, which will announce the proximity of the substance by a loud rattle of the counting mechanism. The apparatus may also be useful to the doctor for controlling the possible secretion in the urine of an artificial radiaoctive substance [8]) administered to a patient. The presence of γ-rays, X-rays, ultraviolet rays and visible light can be demonstrated indirectly with the "electron counter", since these radiations free electrons from the cathode of the counting tube (photoelectric effect). In the abovementioned demonstration of radioactivity it will usually be the γ-rays which cause the instrument to register. (This was also the case in taking the photograph of fig. 4). In order to experience no difficulty from daylight while using the tube, it is lacquered black. If a small window is made in the layer of lacquer, interesting effects can be obtained with the instrument. It reacts immediately with an obvious increase in its rattling when a window is opened, since then more ultraviolet radiation, which is very active photoelectrically, enters the room. The lighting of a match is noticed by the instrument at many metres distance, etc. We have already mentioned that cosmic rays cause current impulses in the counting tube. If one does not use a special shield of lead or iron to weaken these rays, the apparatus counts about 13 impulses per minute; with a shield of 10 cm of iron the number falls to 6 per minute.

[7]) It is known that several years ago in the Netherlands such a case occurred, and a preparation lost in an institution in the Hague was found in a rubbish dump in Drente (a distance of 150 km).

[8]) See on this subject the article of F. A. Heijn and A. Bouwers in the foregoing number of this periodical 6, 46, 1941).

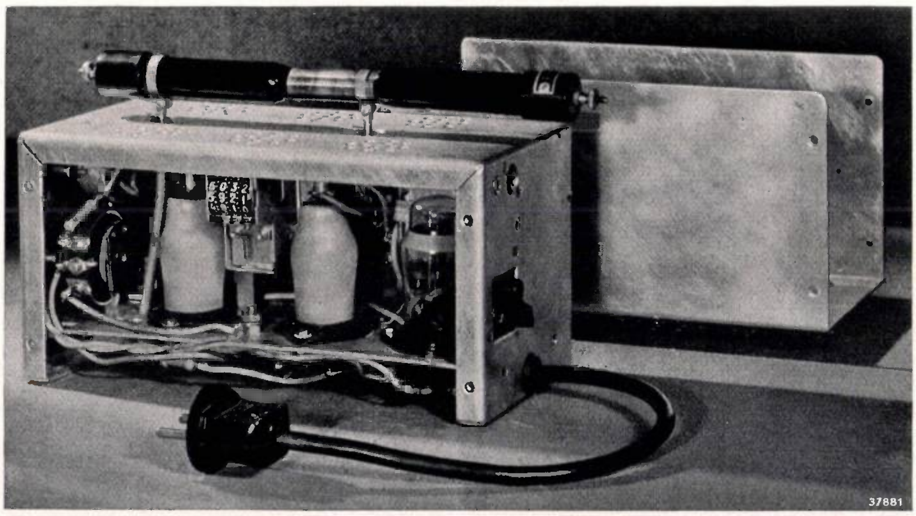


Fig. 7. The complete electron counter opened. On top of the cabinet may be seen the counting tube. Through a small opening in the cover the position of the counting device can be read off. Through two holes on the right side the supply voltage for the counting tube and the grid bias of the first amplifier valve can be varied with a screw driver, if this should be necessary. In the lower part of the right-hand side is the connection for the mains voltage and a switch, above, a connection by which the current impulses can be taken from the valve B_1, and if desired fed to an automatic counting circuit when measurements with a higher resolving power must be carried out. The apparatus is 23 $\times$ 12 $\times$ 12 cm and weighs only $2^1/_2$ kg.

STEREOPHONIC RECORDING ON PHILIPS-MILLER FILM

by K. de BOER.

534.76 : 681.84.081

Beginning with an ordinary apparatus for recording sound film by the Philips-Miller system, an experimental arrangement has been worked out for the making of stereophonic recording. Two sound tracks must be cut in the film strip in this case. The fact that the tolerances as to time differences in the two tracks are extremely small and the requirement that the stereophonic recording must also be able to be played without the stereophonic effect if desired, led to an arrangement in which two sound recorders side by side cut their tracks on the same roll of film. The mechanical problems which were hereby encountered, as well as several details of the arrangement used for recording and reproduction are briefly discussed.

When the time comes to make use of stereophonic reproduction in the cinema, in broadcasting, etc., and the opinion becomes more and more general that the improvement in quality so obtained is worth the trouble, it will become necessary in the first place to find a process of making stereophonic records on a large scale. One possibility of stereophonic recording has already been discussed some time ago in this periodical, namely, recording on gramophone disc records [1]). Although this method was satisfactory for the purpose then in view: the study of different arrangements for stereophonic reproduction, the determination of the requirements to be made of the sound image, etc., for use on a large scale, particularly in the cinema, it can scarcely be considered. The objections are the same as those which formerly defeated the gramophone record in the development of sound film: the playing time of the gramophone record is too short, and there is no easy method of assembly and synchronization with the corresponding picture film. The first objection is particularly important with stereophonic disc records, since the method of recording limits the playing time to half that normally obtained.

Experiments have therefore been carried out in this laboratory in order to work out a method of making stereophonic records on film strips, and the method and apparatus chosen were those of the Philips-Miller system [2]). A short description will here be given of these experiments which led to the solution of several problems which will also be encountered in the construction of a cinema machine with stereophony.

For stereophonic reproduction the sound in the recording room must be taken up by different microphones, while the sound contribution of each microphone is reproduced by a separate loud speaker in the reproduction room. For each sound contribution therefore a separate "channel" is necessary, and in recording, a separate sound track.

[1]) K. de Boer, Experiments with stereophonic records, Philips techn. Rev., 5, 182, 1940.

In practice it is found that two sound contributions suffice to give a satisfactory stereophonic effect [3]); therefore only two sound tracks need be recorded on the film.

How can these two sound tracks be obtained on the film strip?

The ordinary apparatus for recording and reproduction by the Philips-Miller system, see *fig. 2*, is carried out in duplicate, in order, by alternative use of the two halves, to be able to record and reproduce passages of any desired length without interruption. It was logical to use the two recording systems thus available in our experiments for cutting the two sound tracks, namely by drawing the film from the feed roll in one system along both

[2]) The principle of this system is shown in *fig. 1.* On a strip of celluloid film *C*, a gelatine layer *G* with a very thin non-transparent covering layer *D* is deposited. This "Philimil" film is pulled along under a wedge shaped cutter *S*, which is moved up and down in the rhythm of the sound vibrations and thereby cuts a transparent track of varying width (the sound track) in the film. The reproduction of this track, like that of photographically recorded sound film, takes place by an optical method. The particulars of the system are discussed in detail in Philips techn. Rev. 1, 107, 135, 211, 230, 1936; 5, 74, 1940. On the sound recorder see especially the second article referred to (A. Th. van Urk).

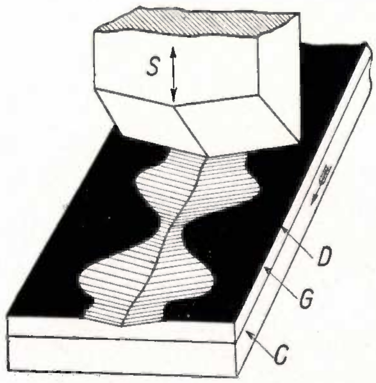


Fig. 1. Principle of sound recording by the Philips-Miller system.

[3]) K. de Boer, Stereophonic sound reproduction, Philips techn. Rev. 5, 107, 1940; Diss. Delft 1940.

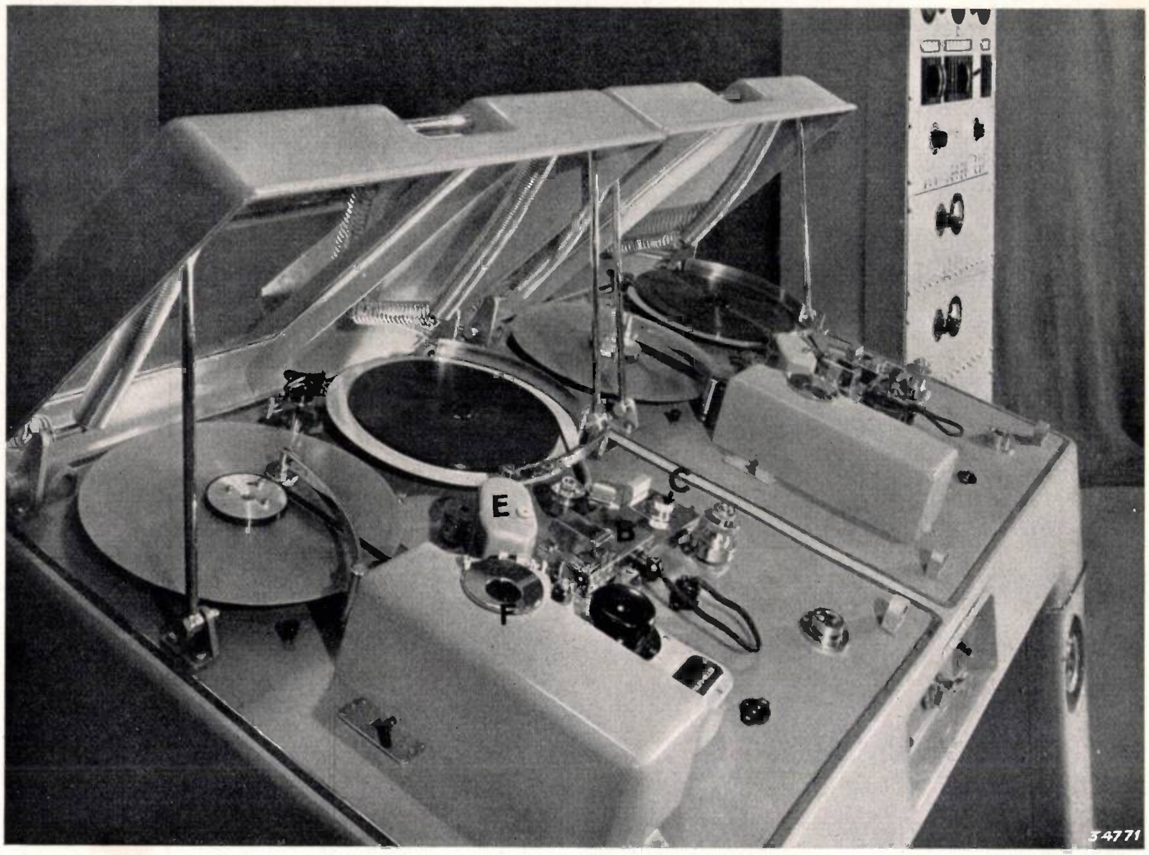


Fig. 2. Apparatus for sound recording by the Philips-Miller system. The machine consists of two similar halves. On the top of each half in the middle may be seen the sound recorder *B*. In *E* and *F* are a lamp and a photocell for reproducing the sound track. To the left and right at the back of each half may be seen the film rolls.

of the sound track cutters to the winding roll in the other system; see the diagram in *fig. 3*. Corresponding points of the two tracks recorded side by side in this way lie at a certain distance *a* (about 1 m) from each other (*fig. 4*). In reproduction the two optical scanning systems, along which the film is moved in the same way as in the recording, must stand at the same relative distance *a* (more precisely, the section of film between the two systems must have the length *a*).

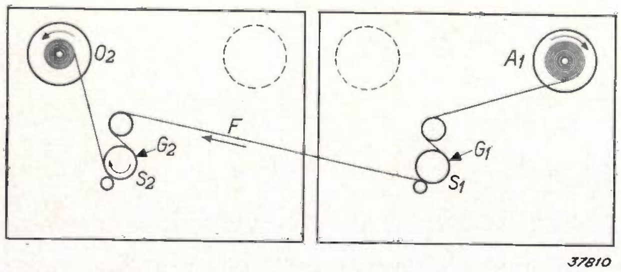


Fig. 3. Diagram showing how stereophonic recordings may be made whereby the two similar halves of the Philips-Miller apparatus are used. A_1 feed roll of film, F "Philimil" film, S_1, S_2 cutting rollers on which at G_1 and G_2 the sound tracks are recorded, O_2 winding roll for film. The cutters G_1 and G_2 are at different heights above the cover plates, so that the two sound tracks lie side by side on the film. The film is transported by the roller S_2.

This apparently simple method encountered various difficulties. In the first place, in recording, a certain stretching of the film occurs due to the tension which must be applied to overcome the resistance by the two cutters (see footnote [2]) in cutting the track in the film. With ordinary recordings this stretch causes only a slight, quite inaudible shift in pitch; in stereophonic reproduction, however, it has unpleasant results. Corresponding points on the two tracks are farther apart due to the stretch, and thus do not pass through the two scanning systems at the same time, but with a slight difference. The result is a displacement of the sound image observed by the listener [4]). In order to avoid this effect the requirement should

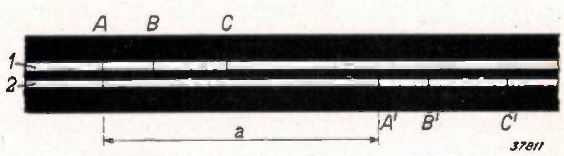


Fig. 4. Diagram of stereophonic sound film recorded according to fig. 3. Corresponding points AA', BB', CC' on the two tracks are separated by a distance a.

[4]) See the article mentioned in footnote [1]).

be made that the stretch of the film over a length of about 1 m (the distance a) should not be more than 60 μ! It must also be noted that a constant stretch over the whole film would be harmless, since one could then eliminate the time difference by changing the relative positions of the scanning systems correspondingly. However, the stretch is not constant along the whole film, since the resistance of the strip, and therefore also the necessary tension, depends upon the cutting depth of the cutter, which varies widely for loud and soft passages in the sound. The result of the stretch is therefore irregularly varying time differences between the two sound contributions, and thus a fluctuation of the sound image.

Attempts may be made to avoid stretch by giving the film strip a separate transport mechanism at each of the two sound cutters, so that the section of the film between need not transmit any force; see the diagram of *fig. 5*. The two transport mechanisms must then be coupled with each other in order to make sure that they both transport equal lengths of film in equal times. Such a coupling, however, in general leaves the possibility open for slight periodical variations in the velocity of the film at the position of the cutters. This introduces periodic time differences between the two sound contributions which in turn cause displacements of the sound image. Experiments with such an arrangement showed that it would only be possible to limit this effect sufficiently with fairly elaborate and expensive constructions.

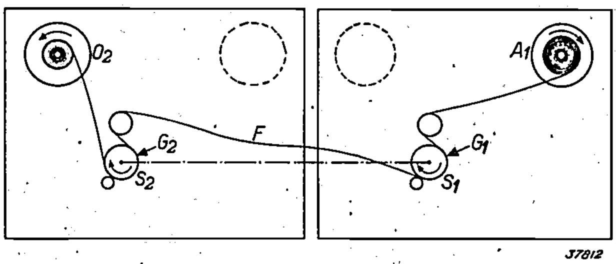


Fig. 5. In order to avoid stretching the film the method of fig. 3 can so be changed that both cutting rollers S_1 and S_2 are driven separately. The two rollers must in this case be rigidly coupled. The letters have the same significance as in fig. 3.

Even if these difficulties were successfully overcome, however, an important objection to the method outlined would remain. As long as the stereophonic sound film has not yet gained a permanent position in the film industry, it is very desirable for stimulating initiative in this direction that the stereophonic film, even without stereophonic effect, should be able to be played off on ordinary sound film machines as well. This is impossible in the method described, since in scanning in the ordinary way the parts of the two sound tracks lying side by side are always made audible at the same time, while they should be reproduced with a time difference of about 3 sec (namely a/v where v = film velocity = 30 cm/sec). The result would thus be that every passage of the recorded speech or music would be made audible twice, at a time interval of 3 sec.

In order to avoid this echo effect the stereophonic film must be recorded so that corresponding points of the two sound tracks lie side by side on the film with a relative displacement of not more than 1 to 2 mm. The problems of stretch and of periodic time differences are thus also eliminated [5]).

These considerations lead to an arrangement in which two sound recorders side by side cut their tracks on the same cutting roller. This idea could not immediately be realized with the existing construction of the Philips-Miller recorder, as is obvious from a consideration of *fig. 6*. It may be seen how on the one hand film and cutting roller, and on the other cutter, armature and magnetic circuit of the recorder are arranged, in relation to each other. If one imagines a second sound recorder, as it were a mirror image of the first, introduced on the other side of the roller, so that the two recorders are side by side, one comes into conflict with the law that two different things cannot be situated at the same time at the same point in space. The most difficult problem arose due to the axis of the cutting roller. As may be seen in fig. 6*a* it would run straight through a pole piece of the second sound recorder. To avoid this, either the distance between axis and cutter, thus the diameter of the cutting roller had to be increased, or the width of the pole pieces of the second sound recorder decreased. The latter is not immediately possible, since the magnetic resistance of the pole pieces may not be made too large. The first solution also has harmful results: the peripheral velocity of the

[5]) It might be thought that the "echo" could also be avoided by only reproducing one of the tracks and covering up the second track in the machine for playing. This would, however, give an unbalanced reproduction of the recorded music, since in stereophonic reproduction the two microphones are expressly so placed that each microphone "hears" the sources of sound in a certain part of the room less well than the other. In order to obtain a satisfactory whole the two sound contributions must in any case, either with or without stereophonic effect, be recombined. Where in this article the term "ordinary" sound film machine is used, it must be kept in mind that "Philimil" film can only be reproduced directly in the corresponding machine, while for the reproduction in an ordinary cinema machine the sound track must first be transferred to the picture film.

cutting roller, *i.e.* the film speed, is fixed; a larger roller must therefore turn more slowly. Because of this the energy content of the flywheel mounted on the axle of the roller, which serves to eliminate the influence of variations in the cutting force, is decreased, while the couple which tends to cause irregularities is made larger due to the larger diameter of the cutting roller (longer lever). Fortunately it was found that in the existing construction the pole pieces of the recorder as well as the flywheel of the drive were of sufficiently generous proportions to make it possible to find a satisfactory compromise. The width of the pole pieces needed only to be reduced by a few mm, while the diameter of the cutting roller could be increased from 35 mm to 70 mm.

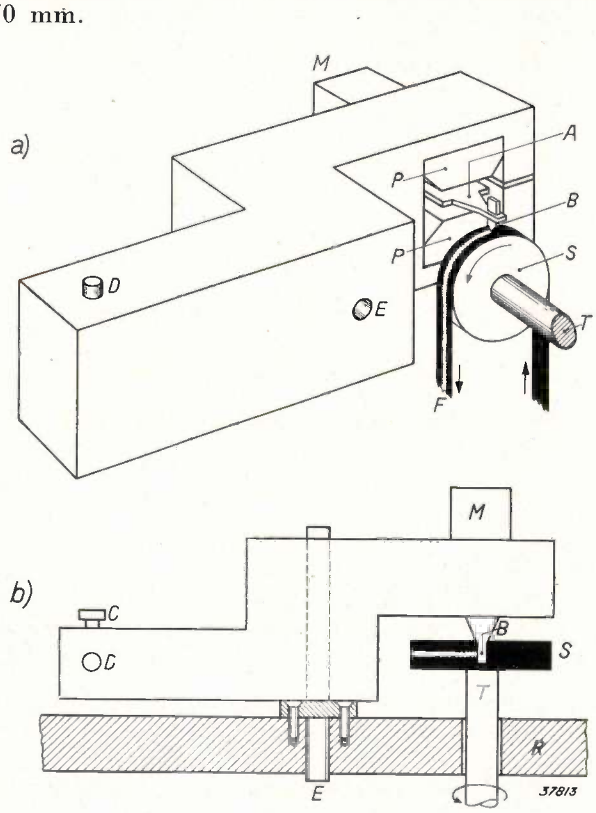


Fig. 6. The sound recorder of the Philips-Miller system. In *a* the relative position of armature *A*, cutter *B*, film *F* and cutting roller *S* may be seen, while *b* shows the assembly on cover plate *R* of the machine. *P* pole pieces, *M* permanent magnet, *T* axis of the cutting roller (also called sound or tone axis), *E* axis about which the whole recorder can be rotated, *D* pin for regulating the cutting depth. The screw for regulating *D* is moved by the knob *C*.

Other less essentiel structural changes were necessary for the body of the recorder and for the magnet which excites the magnetic circuit of the recorder. In the original construction the cutting roller is set into the body of the recorder (see fig. 6), which is a convenient method of giving firm bearings to the shaft *E*, and thereby of fixing as well as possible the relative positions of cutter and cutting roller. Now, in order to be able to place the two recorders side by side, they were made flat, see *fig. 7*. As to the magnet, in the original construction it was set above the pole piece as may be seen in fig. 6.

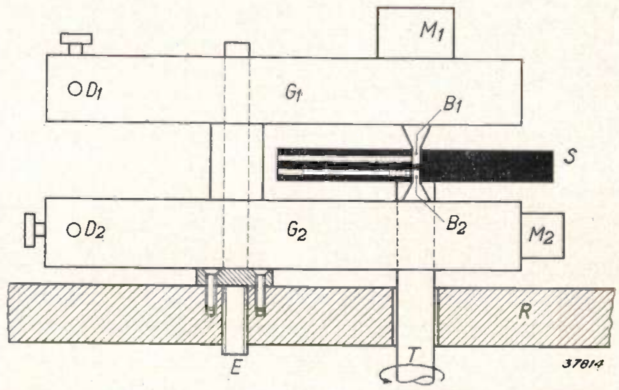


Fig. 7. Construction and arrangement of the two sound recorders G_1, G_2 for cutting two tracks on the same cutting roller *S*. The recorders are on the common axis *E*. Letters have the same significance as in fig. 6.

In the second recorder, which must be as a mirror image to the first, the magnet would fall below the pole piece, *i.e.* between the latter and the cover of the machine. The result would be that the distance between the cutting roller and the cover, which is already greater due to the insertion of the second recorder, would have to be further increased by the height of the magnet. The point of application of the cutting force which acts obliquely on the axle of the cutting roller would thereby lie too far away from the bearings of this axle. A solution was found here by setting the magnet on the pole piece at the side, whereby the course of the lines of force remains practically the same in pole pieces and armature. It is true that the magnet is hereby placed in a position where the space is very much limited by the guiding rollers for the film and by other parts; by using the newest type of magnet steel, however, the dimensions of the magnet, compared with those of the former construction, could be reduced so that no difficulty was experienced.

The final construction used in our experiments is shown in the photographs of *fig. 8* and *9*.

In order if desired to be able to play off the stereophonic film on an ordinary machine, another condition must be satisfied: the total width of the two sound tracks must not be greater than 2 mm. To achieve this purpose each recorder in the arrangement described was provided with a cutter which has the form of half a wedge, see *fig. 10*. The double track obtained in this way, an example of which is reproduced in *fig. 11*, corresponds entirely to a track recorded in the normal way except for a narrow black stripe along the centre and the essential time and intensity differences of the two halves [6]).

Certain precautions had to be taken for the precise adjustment of the cutters. Both cutters must have the same cutting depth in order to obtain the same depth of modulation in the two halves

[6]) In playing on an ordinary machine the two sound contributions are as it were "mixed" in the optical scanning system, which practically comes down to the same thing as the usual mixing of the contributions of different microphones by an electrical method.

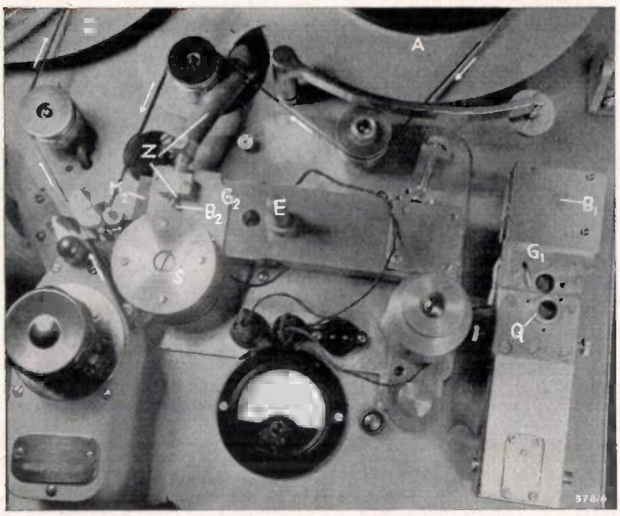


Fig. 8. The lower sound recorder G_2 ready for use. The upper sound recorder G_1 lies at the side. B_1, B_2 are the cutters, S the cutting roller; F the "Philimil" film whose path from the roll A to the roll O is indicated by arrows; E the axis which bears the two recorders; Z mouthpiece and tube of a suction arrangement for removing the shaving cut from the film; M_2 magnet, Q rotating ring with excentric bore for the axis E.

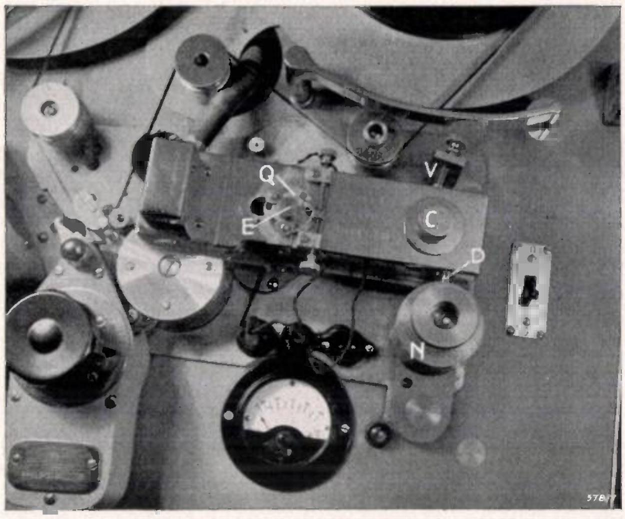


Fig. 9. Like fig. 8 with the upper sound recorder in position. With the cam N (excentric drum) both the recorders are turned about the axis E at the same time against the springs V so that the cutters begin to work.

of the double track. The cutting depth is regulated separately for each recorder by screwing out the pin D visible in figs. 6 and 9 more or less with a micrometer screw C. At the beginning or end of the recording both cutters are set on the film or taken off it simultaneously by means of a cam which pushes the two pins D away at the same moment and thereby turns the recorders about the common axis E; see fig. 9.

A second adjustment relates to the relative distance between the cutters in the direction of length of the film. We have stated above that this distance may amount to a maximum of 1 or 2 mm for the playing of the resulting film on an ordinary machine (without stereophonic effect). For playing with stereophonic effect, where the two sound tracks are reproduced separately, the restriction in this respect is very much greater. A time difference between the two sound contributions of only 2×10^{-4} sec is already observed as a displacement of the sound image [4]). This time difference corresponds to a difference in position on the film of $v \times 2 \times 10^{-4} = 0.06$ mm. (From this the abovementioned tolerance for the stretch is also derived).

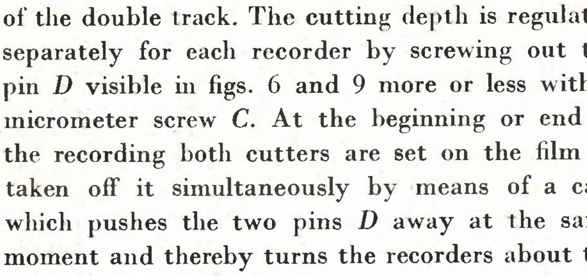
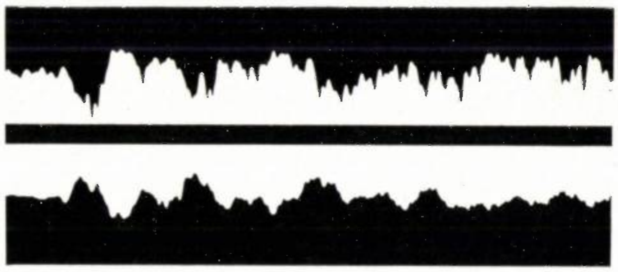


Fig. 10. For cutting the two sound tracks two sapphire cutters are used (b), which are half wedge-shaped, in contrast to the complete wedge-shape of the ordinary cutter (a). The width d of each cutter remains the same in order to be able to clamp the cutter sufficiently firmly to the armature of the sound cutter. The angle of the wedge is actually much more obtuse that here shown (87°).



37861

Fig. 11. Section of a stereophonic sound film recorded by the method described. The two sound tracks in which the two sound contributions are recorded are separated by a narrow black stripe. The differences may be seen in intensity and time of the two vibration forms whereby the stereophonic effect is obtained. The intensity differences are greatest for the high frequencies: the latter are much stronger in the upper track than in the lower.

In order to be able to adjust the position of the cutters with such precision the shaft E in one recorder is excentrically borne by a rotating ring Q, see fig. 8. By a slight rotation of the ring with the help of the screws visible in fig. 9, the recorder with cutter and all is displaced slightly in the direction of length of the film; the adjustment can be carried out by this means with a precision of less than 0.02 mm [7]).

[7]) This method corresponds to that which was formerly used in this laboratory for placing the holes in the Nipkow disc. See H. Rinia and L. Leblans, Philips techn. Rev. 4, 42, 1939, especially page 47.

For stereophonic reproduction the film may be illuminated in the ordinary way through a narrow slit, while the light transmitted by each

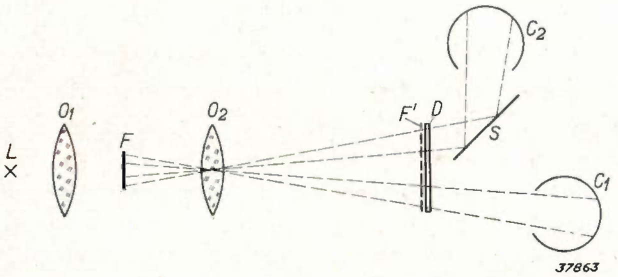


Fig. 12. Arrangement for the reproduction of stereophonic film. With the light source L and the lens O_1 the film F (which is drawn toward the observer perpendicularly to the plane of the drawing) is illuminated, while the lens O_2 projects a sevenfold enlarged image F' of the film on the slit D. Half of the beam transmitted through this slit, corresponding to one of the two sound tracks, is directed to the photocell C_2 with the help of the mirror S, while the other half of the beam is received by the photocell C_1. The alternating currents given by the two photocells are amplified separately and fed to two loud speakers.

half of the double track must be received in a separate photocell. This splitting of the transmitted beam, which may be accomplished very easily with the help of a mirror placed in front of one half of the slit, would here, due to the narrowness of the beam (2 mm), lead to a very fine and delicate construction. The following method has therefore been chosen. With the help of an objective a sevenfold enlarged image of the film is projected on a slit. Half of this image is deflected by a mirror. The arrangement may be seen in *figs. 12* and *13*.

In conclusion we wish to point out that the ordinary "Philimil" film is wide enough (7 mm) to contain, beside the normal or double track of 2 mm, still another sound track. This possibility may be of importance when, as is at present often the case in the cinema, it is desired to record and reproduce the low and high tones separately. As has already been mentioned [3]) the low tones (below 300 c/s) contribute little or nothing to the stereophonic effect. The corresponding sound track therefore need only be a single one, and the position of points on this track with respect to corresponding points on the other (double) track need to be carefully determined. The simple method outlined in the beginning, of recording at differents spots, might therefore be used here, since it would in any case be difficult to record three tracks at the same spot side by side. This extension of the method can, however, only come under consideration when the requirement that stereophonic films must be able to be played on "ordinary" machines will have been outgrown.

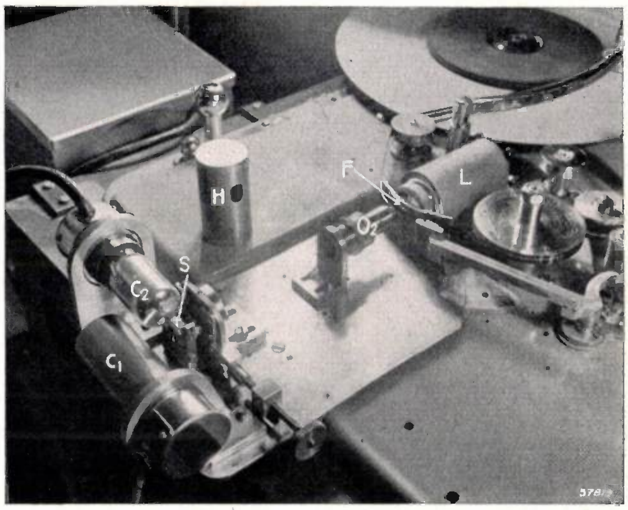


Fig. 13. Photograph of the arrangement for playing stereophonic film. Letters have the same significance as in fig. 12. The cover H of the photocell C_2 has been removed.

SWITCHBOARD WIRE FOR TELEPHONE INSTALLATIONS

by H. FEINER and J. HOEKSTRA.

621.315.616

Switchboard wire for telephone installations is usually insulated with textiles which are provided with an impregnation or a layer of lacquer for protection. In this article the improvement in the insulation is discussed which could be attained by a special choice of raw materials. Especial attention is paid to chlor-rubber lacquer which combines excellent dielectric properties with the advantage of being non-combustible. In conclusion, the technical use of switchboard wire is discussed.

The technical installations in modern telephone exchanges and repeater stations are built up of bays which contain numerous connections and are connected to each other and to the cores of the telephone cable (cable distributor bay).

Requirements are made of these connections, which almost always consist of pairs, which in general correspond to the requirements made of telephone cables. In the first place there is the requirement of adequate insulation. Furthermore, For stereophonic reproduction the film may be illuminated in the ordinary way through a narrow slit, while the light transmitted by each

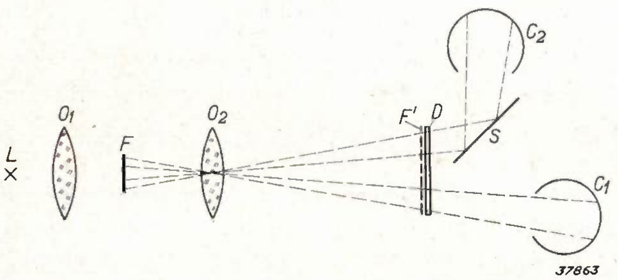


Fig. 12. Arrangement for the reproduction of stereophonic film. With the light source L and the lens O_1 the film F (which is drawn toward the observer perpendicularly to the plane of the drawing) is illuminated, while the lens O_2 projects a sevenfold enlarged image F' of the film on the slit D. Half of the beam transmitted through this slit, corresponding to one of the two sound tracks, is directed to the photocell C_2 with the help of the mirror S, while the other half of the beam is received by the photocell C_1. The alternating currents given by the two photocells are amplified separately and fed to two loud speakers.

half of the double track must be received in a separate photocell. This splitting of the transmitted beam, which may be accomplished very easily with the help of a mirror placed in front of one half of the slit, would here, due to the narrowness of the beam (2 mm), lead to a very fine and delicate construction. The following method has therefore been chosen. With the help of an objective a sevenfold enlarged image of the film is projected on a slit. Half of this image is deflected by a mirror. The arrangement may be seen in *figs. 12* and *13*.

In conclusion we wish to point out that the ordinary "Philimil" film is wide enough (7 mm) to contain, beside the normal or double track of 2 mm, still another sound track. This possibility may be of importance when, as is at present often the case in the cinema, it is desired to record and reproduce the low and high tones separately. As has already been mentioned [3]) the low tones (below 300 c/s) contribute little or nothing to the stereophonic effect. The corresponding sound track therefore need only be a single one, and the position of points on this track with respect to corresponding points on the other (double) track need to be carefully determined. The simple method outlined in the beginning, of recording at differents spots, might therefore be used here, since it would in any case be difficult to record three tracks at the same spot side by side. This extension of the method can, however, only come under consideration when the requirement that stereophonic films must be able to be played on "ordinary" machines will have been outgrown.

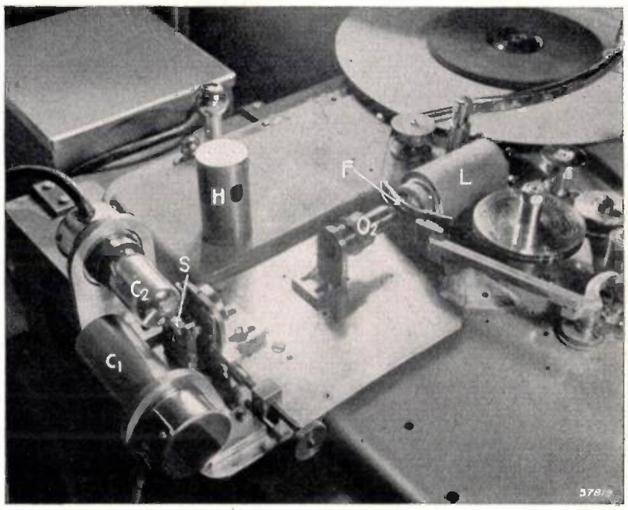


Fig. 13. Photograph of the arrangement for playing stereophonic film. Letters have the same significance as in fig. 12. The cover H of the photocell C_2 has been removed.

SWITCHBOARD WIRE FOR TELEPHONE INSTALLATIONS

by H. FEINER and J. HOEKSTRA.

621.315.616

Switchboard wire for telephone installations is usually insulated with textiles which are provided with an impregnation or a layer of lacquer for protection. In this article the improvement in the insulation is discussed which could be attained by a special choice of raw materials. Especial attention is paid to chlor-rubber lacquer which combines excellent dielectric properties with the advantage of being non-combustible. In conclusion, the technical use of switchboard wire is discussed.

The technical installations in modern telephone exchanges and repeater stations are built up of bays which contain numerous connections and are connected to each other and to the cores of the telephone cable (cable distributor bay).

Requirements are made of these connections, which almost always consist of pairs, which in general correspond to the requirements made of telephone cables. In the first place there is the requirement of adequate insulation. Furthermore, both in the case of switchboard wire and in that of a telephone cable, a compromise must be sought between the total cross section of the insulated double wire and the capacity per unit length. The thinner the layer of insulation, the smaller the cross section and the higher the capacity. With telephone cables many kilometres in length the capacity between the wires with A.C. voltage amounts in a certain sense to a shunt for the apparatus connected, which, even for the relatively low frequencies of speech, may lead to considerable voltage losses. In order to keep these losses as low as possible an upper limit of about 30 pF/m is prescribed for this capacity. In the case of the telephone switchboard wires which have a maximum length per connection of 20 to 30 metres, a capacity of 100 tot 130 pF/m can be allowed. The fact that a decrease in the cross sections can hereby be obtained is a great advantage in connection with economical assembly in the bays, where within a relatively small space a very large number of connections must be carried out.

This accumulation of connecting wires forms an extremely complicated whole during the assembly (*fig. 1*), and one is therefore compelled to mark the wires. The usual method of doing this is to give the wires different striking colours or colour combinations. It is therefore not by chance that a bunch of telephone wires always presents such a pleasant appearance, aesthetics is here the daughter of technology.

Fig. 1. Above: the wires of a telephone bay are laid out on an assembly table and bundled. Below: the result obtained. In the photograph, which does not show the colours, the sorting out of the wires would appear to be hopeless.

The requirement must therefore be made of the layer of insulation that it can be prepared in different colours. In order that the colours shall not become dimmed after some time by dust and dirt it is also very desirable that the surface should be smooth.

Finally it has become more and more desirable lately that telephone switchboard wire should not be inflammable, at least that it should not burn further when it has once been ignited. This requirement should be strictly enforced for the sake of uninterrupted telephone service, since it has already been demonstrated more than once what an enormous disturbance can be caused in telephone communication by a fire, which is in itself not serious, but which attacks the bunches of switchboard wire.

Summarizing we may therefore formulate the following requirements for the insulation of a telephone switchboard wire:

1) high specific resistance,
2) low dielectric constant,
3) smooth surface with bright colours,
4) non-inflammability.

When the wire is used in installations for carrier-wave telephony, in which alternating currents of much higher frequency occur than in normal low-frequency telephony, the value of the capacity is of much more importance, while moreover attention must also be paid to the dielectric losses of the insulation material. The ordinary insulation materials used for telephone switchboard wire cannot therefore be used for wiring carrier-wave systems. In this article we shall first discuss the construction of the insulation of a mechanically strong, non-inflammable switchboard wire for low-frequency telephony, and then show how it has been possible to improve the dielectric properties of the layer of insulation to such an extent that a wire is obtained which is also suitable for carrier-wave systems.

Telephone switchboard wire for low-frequency telephony

The customary insulation of telephone switchboard wire consists of textiles which are impregnated in order to obtain a more compact whole. Cotton is spun or braided around the wire and impregnated with wax, paraffin, ceresin or a mixture of these substances. The insulations thus obtained are satisfactory from an electrical point of view in a dry atmosphere; the surface, however, is not sufficiently smooth, and these wires are very inflammable.

The inflammability disappears when insulation waxes containing chlorine are used as impregnating agent (chlorinated naphtaline for example). Upon heating, these substances develop gases which have a smothering effect and prevent the insulation of the wire from bursting into flame. The waxes containing chlorine have, however, the practical disadvantage that when the impregnating agent must be heated during manufacture very poisonous gases are developed, while the solid insulation waxes may sometimes also have toxic effects through the skin. This objection has been successfully met by using chlor-rubber instead of chlorine waxes. Chlor-rubber is one of the so-called high polymers, which are organic compounds with very long chain molecules and which are too viscous to be used as impregnating agents by themselves. Softeners are, however, known for such substances, usually liquids with low vapour pressures which can be mixed with the high polymers (or resins) in question and give them greater elasticity. With a non-inflammable softener chor-rubber gives a very fire-resistant impregnating agent, which in combination with normal textile can be successfully used in all cases where the electrical requirements are not too high. *Fig. 2* shows a piece of telephone switchboard wire prepared in this way.

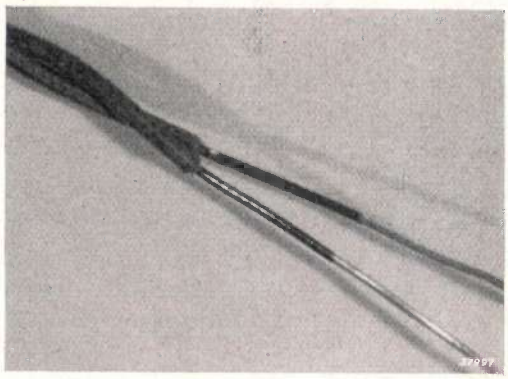


Fig. 2. Cross connection wire consisting of a tinned copper core surrounded by a layer of enamel lacquer and a layer of impregnated cotton.

If a smoother surface is desired than is obtained in this way, the surface of the textile can be covered with a layer of lacquer. Such a layer of lacquer is found to protect the textile insulation just as well as impregnation, so that impregnation is unnecessary. At the same time the dielectric constant is lowered. The usual lacquers for this purpose, which do not, however, satisfy the requirement of non-inflammability, have cellulose derivatives as bases, usually cellulose acetate, a substance which is also used for making artificial silk (acetate silk). The lacquered wire is becoming more and more popular since it is better than the impregnated wire in all the properties mentioned until now.

The influence of moisture

One general disadvantage of the insulation agents described is that the protection against the influence of moisture is only temporary. This is true for both the impregnated and the laquered wire. It is found that the absorption of moisture by the textile is slowed up by covering it with a layer of lacquer, but that it is quantitatively scarcely diminished at all. Protection against the influence of moisture is thus only obtained when this influence is of short duration. Even when materials having a hydrocarbon character are used for impregnation or lacquering, which materials are particularly impenetrable for water, after several days, the textile fibre nevertheless reaches an equilibrium with the moisture content of the surrounding atmosphere. Because of this, for example, with dry cotton in an atmosphere with a relative humidity of 90 per cent, which is quite common near the sea, the dielectric constant may increase by a factor 1.5, the dielectric phase displacement by a factor 5 and the conductivity by a factor 10^6.

It is chiefly the great increase in the dielectric losses in moist air which makes the types of wire discussed unsuitable for carrier systems. In the case of low-frequency telephony on the other hand, only the increase in conductivity is disturbing. This is manifested with a D.C. voltage by corrosion phenomena in the first place. The copper core is electrolytically attacked by the moisture so that ions enter the moisture and thus spoil the insulating property. This process has been completely suppressed by providing the switchboard wires with a layer of enamel laquer *under* the textile. As has been explained in a previous article in this periodical [1]) such a layer is practically insensitive to moisture as far as its insulation capacity is concerned. The increase in the conductivity due to the action of atmospheric moisture is thus practically eliminated. The influence of the moisture on the dielectric properties of course remains.

In the development of a switchboard wire for carrier systems the combatting of the influence of moisture forms the chief problem [2]). In addition there is also the problem of making the lacquered wire non-inflammable as has been done for the impregnated wire.

Switchboard wire for carrier systems

In the development of the switchboard wire for carrier systems the composition of the insulation just described is retained in principle: a tinned [3]) copper wire is covered with a layer of enamel around which is textile and then a layer of lacquer. By suitable choice of textile and lacquer the sensitivity to moisture of the wire is reduced as much as possible and the wire made non-inflammable.

As for textile, cotton (cellulose) and viscose (cellulose hydrate) are much too hygroscopic; real silk and various kinds of acetate silk [4]) may, however, be used. Real silk was rapidly superseded since it is several times as expensive as a good acetate silk (diacetate silk) which is just as good as real silk from an electrical point of view (mechanically real silk is better). Triacetate silk is even better than real silk electrically, and its price is about the same.

In order to give some idea of the electrical properties of kinds of textiles the variation of the insulation resistance, the dielectric constant and the dielectric phase displacement as functions of the relative humidity of the air are given for a number of textiles in *figs. 3, 4* and *5* [5]). The difference between cotton and artificial silk is particularly large at

1) J. Hoekstra, Properties and applications of enamelled wire; Philips techn. Rev. 3, 39, 1938.

2) This problem is particularly important where carrier telephony must be applied in tropical or subtropical countries, where the air is often very moist. Carrier telephony may offer an economically advantageous possibility of fulfilling a function which is often necessary in such countries: that of connecting two communities between which a large number of conversations must be conducted simultaneously, and whose distance apart is so great that the extra expense of a cable with many cores would be greater than the extra expense of a carrier-wave apparatus.

3) The copper wire is tinned in order to facilitate soldering. As to the mechanical properties of the enamel layer, an untinned wire would be better, since the enamel adheres less well to tin than to copper. Nevertheless, the tinned and then enamelled wire can be wound around a mandril equal to twice the diameter of the wire without the occurrence of cracks in the enamel layer.

4) Cellulose acetate or acetyl cellulose is a compound of cellulose with one or more molecules of acetic acid per glucose residue with the splitting off of a molecule of H_2O. The cellulose molecule consists of a large number of glucose molecules coupled together; each glucose molecule contains three hydroxyl groups (OH-groups) which give the cellulose a very hygroscopic character. The greater the number of these groups which are esterified, the less hygroscopic the product. The decrease in hygroscopocity means at the same time a decrease in the influence of atmospheric moisture on the dielectric properties. Therefore triacetyl cellulose is better than diacetyl-cellulose. The former, which no longer contains any unsaturated OH-groups, is, however, difficult to work: due to the lack of hydrophilic hydroxyl groups cellulose triacetate is insoluble in technically useful solvents, and therefore unsuitable for spinning. At the present time methods are known of doing so in spite of this fact; the triacetate silk so obtained is, however, considerably more expensive than diacetate silk.

5) The measurements were carried out on wires which had two layers of textile spun around them. The results are of only relative value since they depend on the spinning technique used.

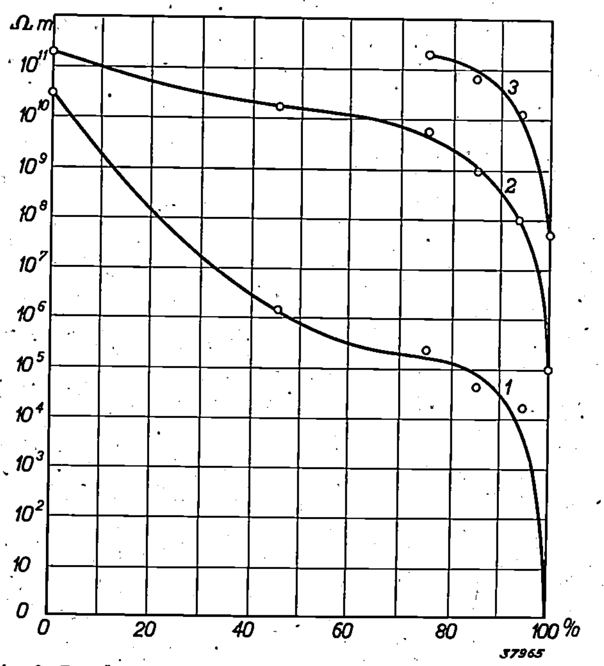


Fig. 3. Insulation resistance of different kinds of textiles as a function of the relative humidity of the air to which the wires were exposed for 48 hours at 25 °C. *1* cotton, *2* diacetate silk, *3* triacetate silk. It is found that the insulation values in the dry state are not very different, but that at high relative humidities cotton is the poorest and triacetate silk the best.

higher degrees of humidity. Triacetate silk is appreciably better than diacetate silk, but the difference in quality is not enough to make up for the much higher price of triacetate silk in all cases. Therefore diacetate silk will generally be used, and triacetate silk only when very strict requirements are made about the insensitivity to moisture.

As stated above acetyl cellulose can also be used for the lacquer layer. However, while the artificial silk made of acetyl cellulose which lies around the wire as a porous mass exhibits satisfactory dielectric properties, the same material in a compact form as a layer of lacquer is much less satisfactory. In the latter case the dielectric constant is much higher and the dielectric losses, which are characteristic of the material, are much more disturbing. Triacetyl cellulose cannot be used instead of diacetyl cellulose because the former is insoluble in technically useful solvents, and therefore unsuitable for a lacquer. Other cellulose derivatives, however, can be used, such as ethyl cellulose and other cellulose esters, as well as nitro-cellulose[6]) and certain cellulose tri-esters. Some of these substances actually offer greater resistance to moisture than does acetyl cellulose. The improvement obtained is not, however, decisive.

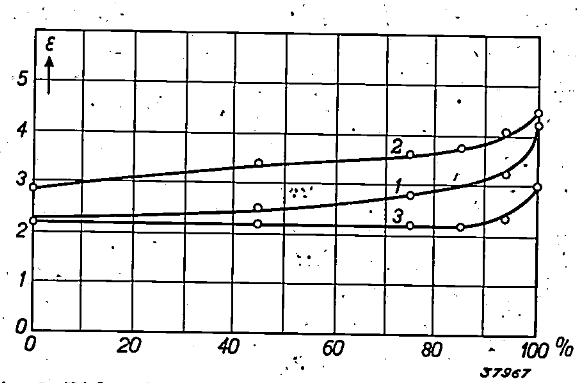


Fig. 5. Dielectric constant of cotton (*1*), diacetate silk (*2*) and triacetate silk (*3*) as a function of the relative humidity.

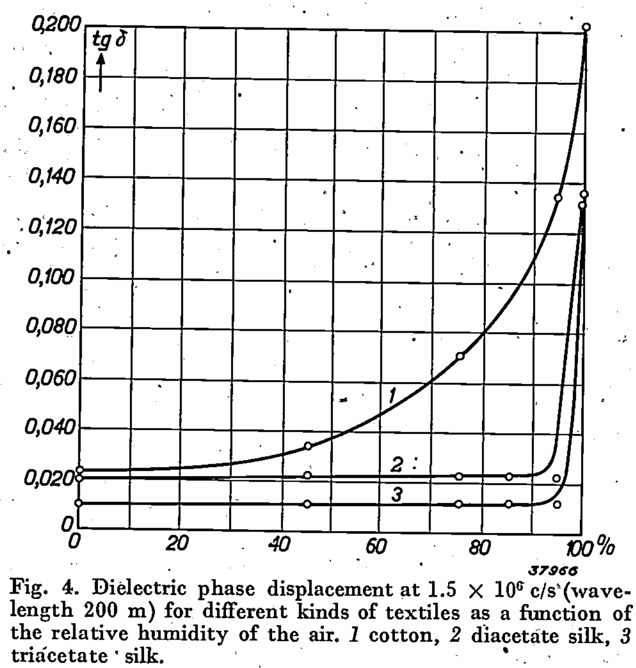


Fig. 4. Dielectric phase displacement at 1.5 × 10^6 c/s (wavelength 200 m) for different kinds of textiles as a function of the relative humidity of the air. *1* cotton, *2* diacetate silk, *3* triacetate silk.

Besides the cellulose products there is still a large number of other laquer substances. These substances, several of which are mentioned in the article referred to above on enamel-laquer wires, are also artificial or natural high polymers, to which resins are sometimes added to increase the hardness, and more often softeners to increase the elasticity. With a suitable choice from among these substances lacquers can be obtained which are more resistant to moisture than the cellulose derivatives and which also possess good mechanical and electrical properties.

We have attempted to combine the possibilities of the high polymers and resins with the requirement of non-inflammability, and have carried out experiments with chlor-rubber which has already been mentioned as an impregnating agent. It was found that from chlor-rubber in combination with a suitable softener a lacquer can be prepared which is far superior to the other lacquer substances mentioned in the foregoing, not only in respect to its non-inflammability, but also in respect to its dielectric properties. It has the highest insulation

[6]) This is, however, not very stable and very inflammable.

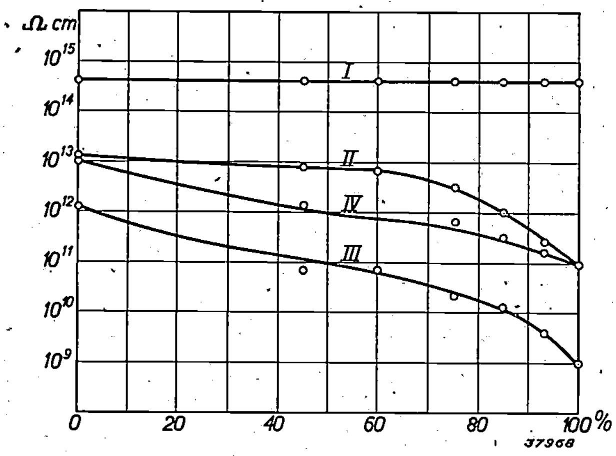


Fig. 6. Insulation resistance of different kinds of lacquer as a function of the relative humidity.
I chlor-rubber lacquer
II ethyl cellulose lacquer
III acetyl cellulose lacquer
IV nitrocellulose lacquer.
It may be seen that chlor-rubber has by far the highest insulation resistance; acetyl cellulose has the smallest insulation resistance, and at the same time it decreases most rapidly with the relative humidity.

resistance, the lowest dielectric constant and the lowest dielectric losses. Moreover, the change in its properties upon a variation in the moisture content between 0 and 100 per cent is almost inappreciable. In *figs.* 6, 7 and 8 the dielectric properties of a number of lacquer substances are given; the superiority of chlor-rubber is very obvious.

As to mechanical properties, chlor-rubber is somewhat less favourable than the other lacquers customarily used in insulation technique. It has been found that this can be remedied by covering the chlor-rubber layer with some kind of cellulose lacquer. Since the electrical field strengths are weakest on the outside, the dielectric properties of the whole combination are made only slightly poorer. In this way a product is obtained which satisfies high requirements electrically as well as mechanically, and which is insensitive to moisture and non-inflammable.

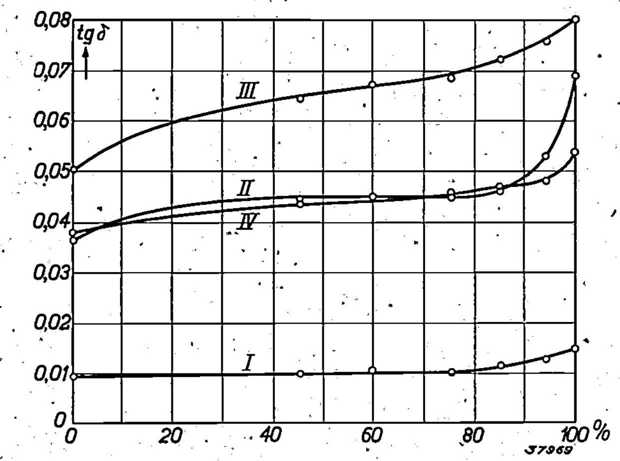


Fig. 7. Dielectric phase displacement at 1.5×10^6 c/s (200 m) for different kinds of lacquer as a function of the relative humidity. The Roman numerals have the same significance as in fig. 6. Chlor-rubber again exhibits the most favourable behaviour.

The manufacture of switchboard wire for carrier systems

The switchboard wire described consists of a tinned copper core, surrounded by four layers of insulation: a layer of enamel, a layer of acetate silk, a layer of chlor-rubber lacquer and a layer of cellulose lacquer. The manufacture of this wire is briefly as follows.

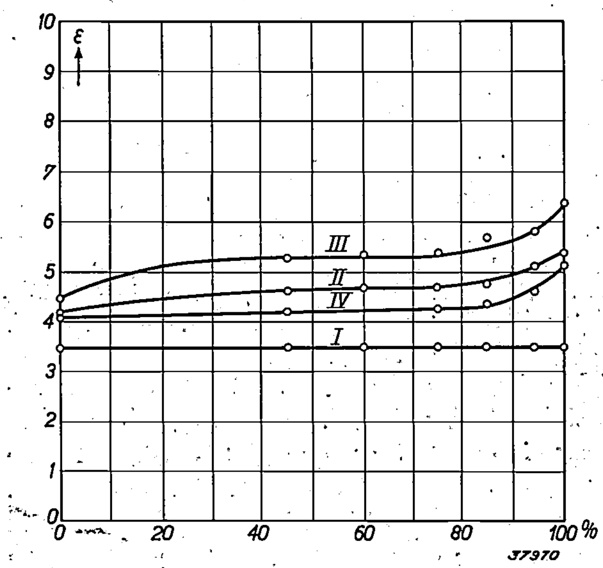


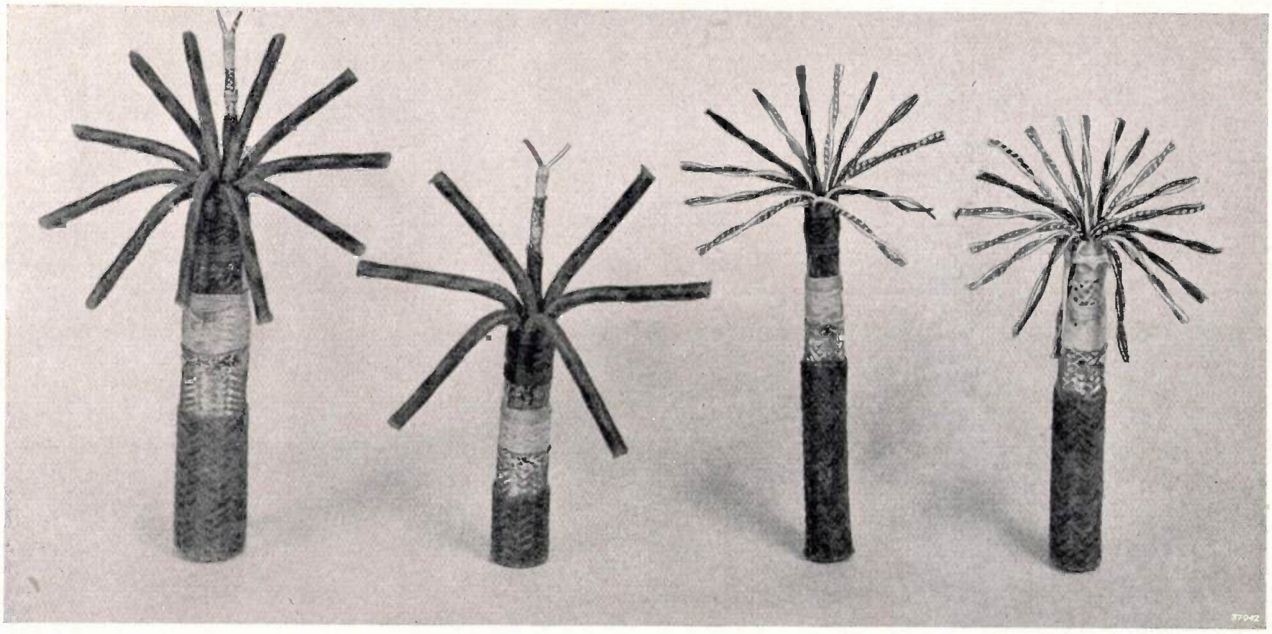
Fig. 8. Dielectric constant at 1.5×10^6 c/s for different kinds of lacquer. The numerals have the same significances as in figs. 6 and 7. Chlor-rubber is again the most favourable.

Tinned copper cores, usually of 0.6 or 0.8 mm diameter are heavily enamelled. The tinning of these cores demands particular care since the tin layer is heated above its melting point when the wire is enamelled, and therefore every point where there is an excess of tin leads to disturbing irregularities in the enamel layer. The thickness of the enamel layer is about 0.06 mm, *i.e.* about 1.5 times the usual thickness. The number of spots where defects in the insulation occur is considerably reduced by this increase in the thickness. If, for example, the number of spots is investigated at which the wire gives a breakdown at 120 volts when it is drawn through a vessel of mercury, a reduction from 1 per 5 m to 1 per 60 m is found.

The artificial silk is now wound around this enamelled core, in four layers for instance. The silk is laid on in loosely twisted strands and in the successive layers it is wound alternately to the left and to the right. The outer layer may to advantage consist of real silk since that is mechanically stronger. This outer layer of textile is made in one or more colours. Over this come the layers of lacquer. These layers, chlor-rubber and over that cellulose lacquer are deposited separately in several extremely thin layers in order to save time in drying, since the sum of the drying times of for instance 10 thin layers is considerably less than that of one layer 10 times as thick. Moreover, such a thick layer of lacquer would not possess the smooth surface which is desired for these wires.

Until now we have spoken only of the manufacture of a single wire. If we now pass on to the manufacture of pairs and the combination of a number of pairs to a cable such as is used in the connections between two bays, quite different problems of construction are encountered. These arise mainly from the requirement that an A.C. voltage acting on a pair may not be transmitted to another pair by capacitative or inductive action. Such couplings give rise to cross talk, *i.e.* the phenomenon in which when an A.C. voltage is applied to one pair at the beginning of the cable a part of that voltage is detected on other pairs as well at the end of the cable. In order to combat these couplings the different pairs are twisted together at a certain pitch before they are twisted in concentric layers into a compact cable. The direction of twisting for the successive layers of the cable, like the pitch of the pairs, is chosen as favourable as possible in connection with the combatting of capacitative and inductive influences, and the pitch of adjacent pairs is different.

When very heavy requirements must be made as



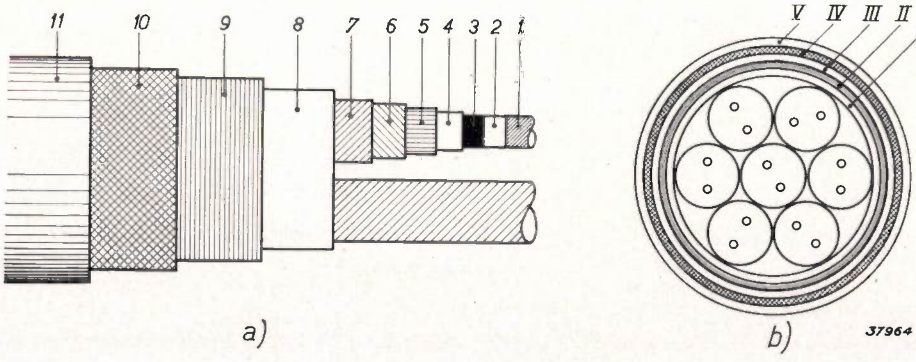


Fig. 9. Cables with shielded and non-shielded pairs. Below, a diagrammatic cross section; *a* of a shielded pair, *b* of a cable with seven shielded pairs.

1 copper core,
2 tinned surface,
3 enamel layer,
4 diacetate silk,
5 coloured real silk,
6 chlor-rubber lacquer,
7 cellulose lacquer,
8 braided layer of artificial silk,
9 film layer,
10 braided covering of tinned copper wire,
11 braided layer of cotton impregnated to be non-inflammable,
I braided cotton,
II film layer,
III layer of spun cotton,
IV braided tinned copper wire,
V braided cotton impregnated to be non-inflammable.

Fig. 10. The bundles of wires of a telephone installation are mounted in the bay.

to the absence of mutual couplings, shielded cables can be used. In the construction of these cables as here described each pair is surrounded by braided tinned copper wire. In *fig. 9* the construction of a shielded and a non-shielded pair may be seen; the text below the figure gives the details. Finally in *table I* a survey is given of the most important technical properties of these cables.

Table I

Properties of shielded and non-shielded cables manufactured by Philips, after tropics test [4]).

Property	non-shielded pairs (0.6 mm Ø)	shielded pairs (0.8 mm Ø)
Working capacity*) in pF/m	max. 100	max. 130
Loss factor*) tan δ at 1 800 c/s.	tg$\delta < 0.1$	tg$\delta < 0.1$
Loss factor*) tan δ at 1 800 c/s before tropics test	tg$\delta < 0.05$	tg$\delta < 0.05$
Insulation resistance in MΩ at 500 V D.C. voltage for a cable of 10 m**)	$>10^4$	$>10^4$
Test voltage**)	2 000 V, 50 c/s, 2 Min.	
Cross talk damping for a cable of 10 m length***)	>90 dB $R = 600\ \Omega$ $f = 4\,000$ c/s	>110 dB $R = 150\ \Omega$ $f = 70\,000$ c/s

*) If desired these values can still be improved by using triacetyl cellulose instead of diacetyl cellulose.

**) In testing the cable one core is connected to the one electrode, all the other cores and the shielding, when present, are connected to the other electrode.

***) The cross talk damping is defined by 20 log E_1/E_2 where E_1 is the voltage which occurs at the end of that pair to whose other end a voltage is applied, and E_2 the voltage which is detected at the end of the cable on any other pair. R is the cut-off resistance of the cross, f the frequency at which the measurement is carried out.

The way in which the switchboard wire is used in shown in fig. 1 as well as *figs. 10* and *11* which show the wiring of a bay in a telephone exchange in three stages of completion. The wiring for the bays is laid out on an assembly table; the wires are joined and bound together as far as possible in bundles. This step was shown in fig. 1.

The finished bundle is now laid in the bay and

[4]) If not expressly stated otherwise the values are determined after a tropics test of three days with an atmosphere of 94% relative humidity and daily temperature cycles with a maximum temperature of 45 °C.

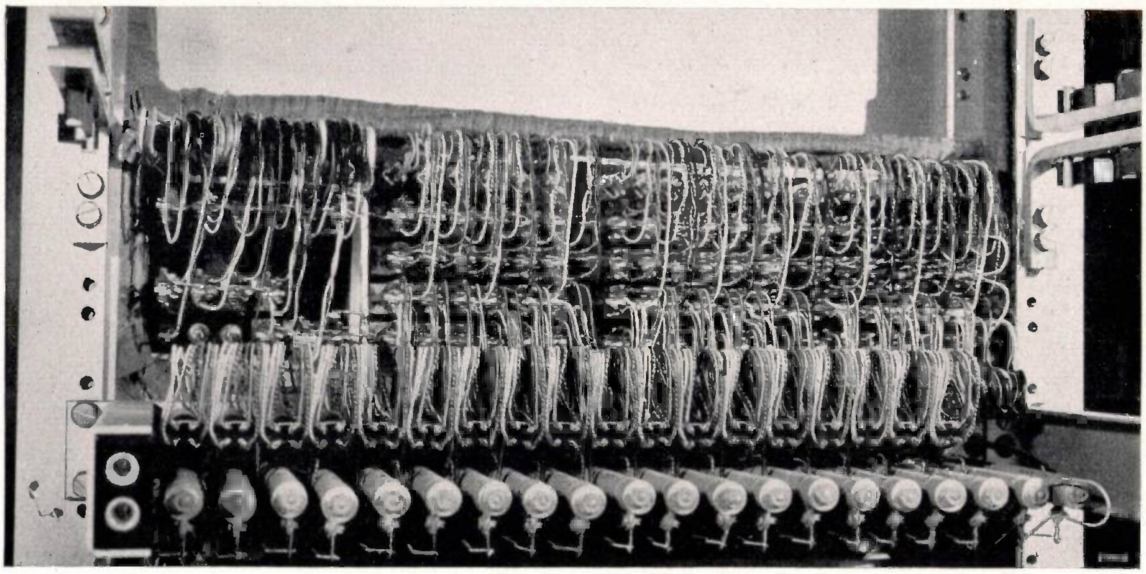


Fig. 11. Part of the wiring of the finished bay.

fastened to it, and the panels with the telephone relays, etc. are mounted. The bay assembly proper then begins (see fig. 10). In fig. 11 a part of the finished bay may be seen.

INVESTIGATION OF THE STRESSES ON PARTS OF A DIESEL ENGINE WITH THE HELP OF THE CATHODE RAY OSCILLOGRAPH

621.317.755 : 620.178.5

Difficulties which arose in the case of an 8-cylinder four-stroke Diesel engine, which gives 600 h.p. at 375 r.p.m., formed the stimulus for the following investigation of stresses [1]). In this engine the bolts which fasten the cap of the main blocks of the crank shaft to the basic metal always broke.

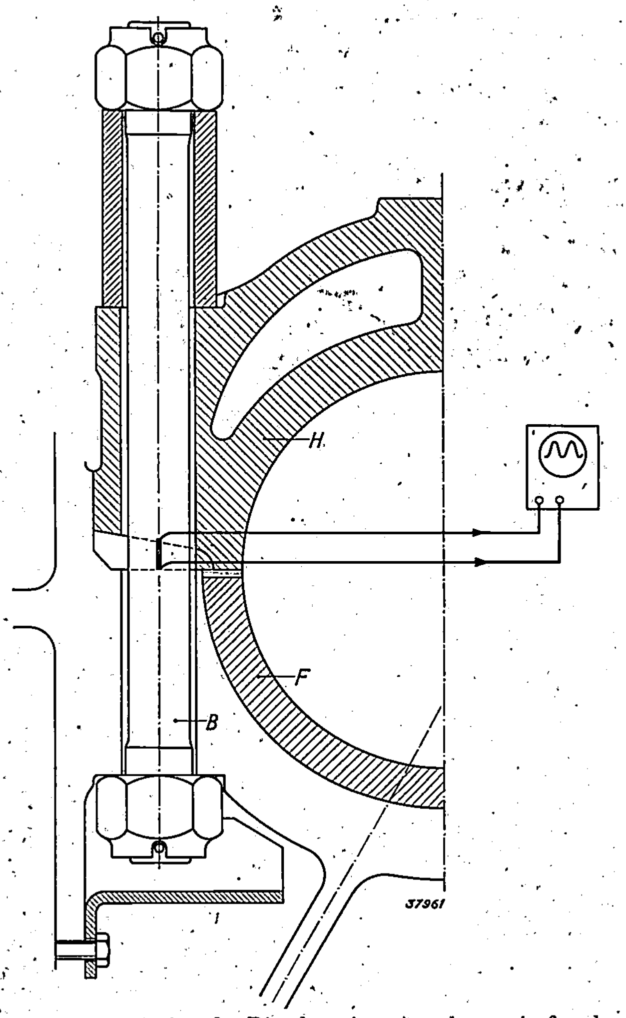


Fig. 1. Main blocks of a Diesel engine. A carbon strip for the measurement of the stress has been placed on the bolt which joins the main blocks with the engine frame.

In *fig. 1* which shows these bolts, the cap H of the bearings may be distinguished, the lower half F of the bearings, which form one piece with the frame, and one of the bolts B which hold the two halves together.

A possible explanation for the breaking of the bolts was that they were too much stretched at the moment when the explosion took place in the cylinder. In order to investigate this, stress measurements were made with the cathode ray oscillograph according to a method previously described in this periodical [2]). This method is based upon the fact that the resistance of a conductor which is obtained by the deposition of a layer of carbon on a strip of elastic insulation material is very much altered by a slight stretching or contraction. A strip of such a carbon layer conductor is glued to the object to be examined so that the carbon layer takes part in any stretching movements of the base. When an electric current of constant magnitude is sent through the resistance a voltage occurs between the extremities which fluctuates due to the stretch and contraction of the underlayer. These voltage variations are made visible with a cathode ray oscillograph. In order to calibrate the oscillogram obtained, the strip is previously subjected to a known stretch by means of a measuring apparatus shown in *fig. 2*.

With a four-stroke engine with 375 r.p.m. the condition in the cylinder changes periodically at a fundamental frequency of 3.1 c/s. The same fundamental frequency may be expected for the varying state of stress in the bolts. This is a much lower frequency than usually occurs in electrical phenomena, and most electrical oscillographs are unsuitable for recording phenomena of such low frequency without deformation of the oscillogram. The oscillograph GM 3 156, which was designed especially for use in tool making [3]), is, however, able to do this, and was indeed used for this investigation.

In fig. 1 the main details of the arrangement may be seen. A number of bolts were provided with resistance strips in the direction of the stretch. The voltage variations occurring were fed, *via* the oscillograph amplifier, to the vertical deflection plates of the cathode ray-tube. An A.C. voltage of the same frequency and phase as the motion of the piston was applied to the horizontal deflection plates.

The picture of the variation of the stretch obtained in this way was found, as was expected, to correspond in the first instance to the pressure variations in the adjacent cylinders of the engine as a function of the time. Two examples are re-

[1]) The data in this article are due to the kindness of Mr. M. J. Visser of the Department for Vibration Testing of the N.V. Werkspoor, Amsterdam.

[2]) Philips techn. Rev., 5, 26, 1940.

[3]) Philips techn. Rev., 5, 277, 1940.

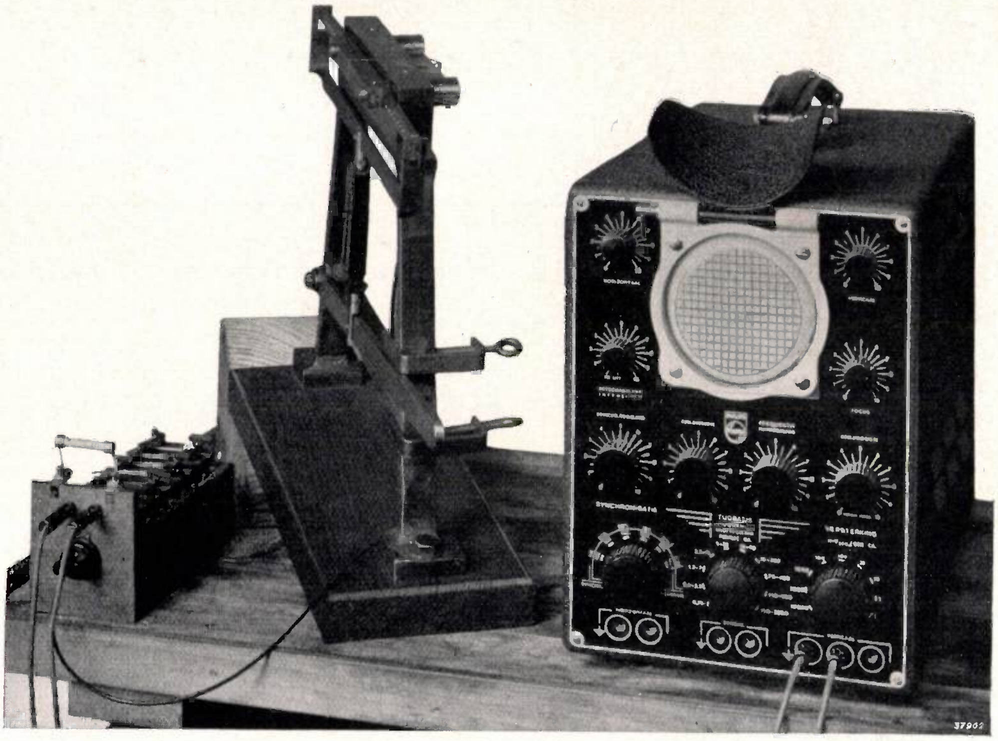


Fig. 2. Arrangement for the calibration of the resistance strips. By means of the setup shown on the left an accurately known bend is given to an elastic spring, upon which the resistance strip glued to it undergoes a known stretch. The change in resistance hereby caused is measured with a bridge connection. The oscillograph apparatus on the right serves for recording the resistance changes in dynamic stretch measurements with the help of the calibrated strip.

produced in *fig. 3.* They refer respectively to a bolt between two cylinders which move in the same rhythm and a bolt between two cylinders whose motion differs 90° in phase. The maximum stretch

a

b

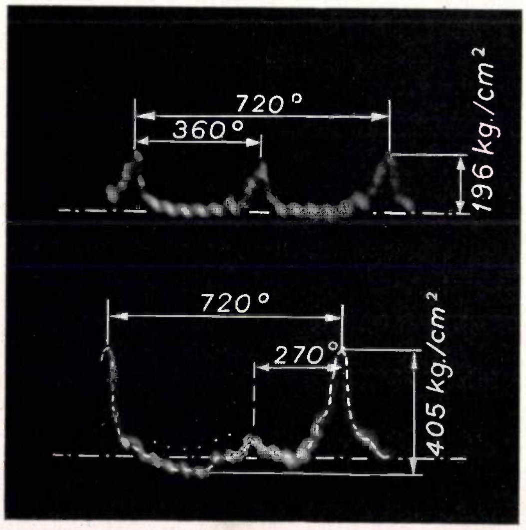


Fig. 3. Oscillograms of the tensile stresses in the bolts of a Diesel engine. *a*) Bolt between two cylinders, which move in the same rhythm; *b*) bolt between two cylinders whose motion differs in phase by 90°.

of the bolts calculated from these diagrams with the help of the calibration curve could in both cases be considered as permissible for the material used.

From the results of the investigation, therefore, it was found that the load on the bolts in normal use could not lead to breakage. It could be concluded from this that during the time when the difficulties were encountered the engine must have been exposed to abnormally high loads. Very probably these loads, which were superposed as extra stresses on the normal ones, loaded the bolts beyond the fatigue limit.

The engine was indeed exposed to very large and sudden changes in load. The result of these shocks was that the basic metals had obtained a certain play on the foundation plate. Because of this, lateral movements could occur upon a too sudden change in the load on the engine, which probably caused a large and irregular stress in the bolts. By a slight change in the construction the lateral motion of the basic metal could be made impossible. After that no further breakage of the bolts occurred.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIKEEN

1516: J. H. de Boer: Atomic distances in small graphite crystals and the nature of the bond (Rec. Trav. chim. Pays Bas **59**, 826-830, July-Aug. 1940).

The decrease in the distance between the carbon atoms in the atom layers of graphite crystals and the increase in the distance between these layers which Hofmann and Wilm found for small graphite particles compared with normal graphite, are an immediate result of the two kinds of binding forces which here play a part. These are the so-called homopolar bonds and the van de Waals-London attracting forces.

1517*: M. J. Druyvesteyn and F. M. Penning: Mechanism of electrical discharges in gases at low pressure (Rev. Mod. Phys. **12**, 87-174, Apr. 1940).

In this monograph a detailed survey is given of the phenomena of electrical discharges in gases at low pressures. The following subjects are dealt with: the distribution of velocity of the electrons; ionization and excitation with a small electric current through a gas; breakdown voltage and Townsend discharge (where the space charge plays practically no part); discharges in which the space charge is essentially important; the glow discharge as well as the arc in the neighbourhood of the cathode; the positive column and finally the phenomena in the neighbourhood of the anode with different forms of discharge.

1518: J. F. Schouten: The residue and the mechanism of hearing (Proc. Ned. Akad. Wet., Amsterdam **43**, 991-999, Oct. 1940).

Continuing with earlier investigations relating to subjective sound analysis and the observation of a fundamental tone which is missing in the Fourier analysis of the sound (see **1372** and **1501**; *cf.* also Philips techn. Rev. 4, 167, 1939), this paper discusses how the perception of a complex sound may be supposed to take place in the human ear (see also Philips techn. Rev. **5**, 286, Oct. 1940). It appears to be permissible to state that the ear performs a Fourier analysis of the sound, at least in so far as its resolving power makes that possible. The earlier investigators were correct when they claimed that every periodic change in air pressure, at least within a certain interval of frequencies, will be observed as a separate tone. If it is desired simply to determine the components of a sound, it is correct to study its Fourier spectrum. If it is desired to determine the sound of every separate component, then one must not consider the frequency of the sound, but the periodicity of the wave form of the vibration by which the corresponding receiver is excited by the complex sound. This periodicity may be considerably lower than that corresponding to the characteristic frequency of the excited receiver. The pitch of the component is determined by this periodicity and not by the characteristic frequency.

1519: E. J. W. Verwey: Absolute grootte van thermodynamische grootheden voor ionen in waterige oplossing en van den electrischen potentiaalsprong aan het vrije oppervlak van water (Absolute magnitude of thermodynamic quantities for ions in aqueous solution and of the electrical potential jump at the free surface of water) Chem. Wbl. 37, 530-535, Oct. 1940).

An estimation is given of the absolute magnitude of the reciprocal energy (change in the free chemical energy) upon the solution of monovalent ions in water. By comparison with the corresponding changes in the total free energies which can be calculated on the basis of a publication of O. Klein and E. Lange, it is found that the potential jump at the free surface between vacuum and water amounts to about —0.5 volt. This negative potential difference means that the water molecules are so arranged that an excess of water molecules have their positive poles directed toward the outside.

*) An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

VOL. 6. No. 4, p. 97-128. APRIL 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

THE MECHANICAL PROPERTIES OF WELDED JOINTS

by P. C. van der WILLIGEN. 621.791.052

The material deposited from a welding rod must as far as possible possess the same mechanical properties as the material of which the pieces to be welded consist. The Philips welding rod type 55 was especially intended for welding the steel St 52; consequently a tensile strength of at least 52 kg/mm² was required of the material deposited in the weld. By a suitable choice of the necessary hardening components the high tensile strength could be obtained without a detrimental increase of hardness in the weld and without diminishing the deformability. The high notched bar toughness of the material deposited is very striking (17 to 20 kgm/cm² for a cross section of the break of 5×10 mm²). This is obtained by a suitable composition of the coating of the welding rod, whereby the nitrogen which depresses the impact value, is prevented from reaching the deposited material. Thanks to the great toughness and deformability (plastic reserve) of the deposited material of the Philips welding rod 55, very hard types of steel, rich in carbon can also be welded, while on the other hand due to the calcium content of the coating it is also possible to weld types of steel of poorer quality contaminated with sulphur. Even free-cutting steel which is very rich in sulphur can be welded with these rods without the occurrence of porosity.

In arc welding the joint between two plates is obtained due to the fact that the heat of the arc fuses not only the welding rod serving as electrode but also a surface layer of the plates, and causes the materials to run together. During this process, in the fused material, on its surface and in the transition zone to the parent metal of the plates, all kinds of chemical reactions take place which may exert a far-reaching effect on the mechanical properties of the finished weld. If certain properties are desired the processes taking place must be controlled in a suitable way. Four main factors may be indicated which are hereby important: the composition of the parent metal, the composition of the welding rod, the environment in which the fused material is situated (coating of the rod) and the rate of cooling.

The first of these factors is not generally under control, since the choice of the parent metal will be determined by other, particularly structural considerations. The second and third factors are the ones which the manufacturer of welding rods can vary in order to adapt his product to the requirements of the constructor. The fourth factor is mainly a question of welding technique.

What are the requirements of the constructor? The mechanical properties of the welded joint must correspond as nearly as possible with those of the parent metal. This generalized formulation of the requirements is by no means obvious. Formerly it was as a rule considered sufficient when the tensile strength of the weld (the material deposited) was the same as that of the parent metal; at present the conclusion has been reached that the other mechanical properties also, such as the elongation, notched bar toughness, etc. should satisfy the same requirements.

From the above it follows that the better the quality of the material to be welded, the higher the requirements which should be made of the welding rod. The opposite is, however, also correct up to a certain point: it is more difficult to obtain a satisfactory weld, the poorer the quality of the parent metal, *i.e.* the more impurities contained in it (sulphur, phosphorus, etc.) which are detrimental to certain mechanical properties (sometimes to the advantage of others). These impurities may for instance cause very unpleasant phenomena upon welding, such as the occurrence of porous spots or cracks in the weld.

A new welding rod has recently been developed by Philips (type 55) with the intention of satisfying different requirements simultaneously. The rod is on the one hand suitable for welding good qualities of steel (St 52), and this in the sense that the material deposited from the rod not only has the same high tensile strength as the steel, but also among other properties an unusually high impact value; on the other hand the welding rod is extremely insensitive to the most dangerous component of the parent metal, namely sulphur. In the following we shall discuss these properties more closely and indicate their relation to the above-mentioned factors under the control of the manufacturer: the choice of composition and coating of the welding rod.

The tensile strength of the weld

In the early days of arc welding only the ordinary kinds of ingot steel, for instance St 37 (*i.e.* a steel with a minimum tensile strength of 37 kg/mm²) were welded. Only about 10 years ago the kinds of steel with a higher tensile strength, among which St 52 is the most important representative, began also to be welded electrically. In Germany several bridges were constructed of welded St 52. The result was at first disappointing: in some of the welded girders cracks appeared, a phenomenon which had never been encountered with St 37. What was the cause?

The difference between the tensile strength of St 37 and St 52 is due to the difference in the contents of so-called hardening elements, carbon among others. If iron containing carbon is cooled rapidly from the high temperature necessary for the preparation of the steel, it is very hard, which also means that it has a high tensile strength and yield value, since the latter are correlated with each other and with the hardness. The greater the speed of cooling, the greater the hardness caused by the carbon.

The process of hardening is now as it were repeated during the welding, and this takes place in different ways at different spots in the weld. The hardening is most pronounced in the transition zone between weld and parent metal, since here part of the latter is heated during the welding to just under its melting point and then very rapidly cooled, due to the good heat conduction of the parent metal. This cooling is considerably more rapid than was the case in the preparation of the material (for instance 50° per second). The result is an increase of hardness in the transition zone. This increase in hardness, which depends further upon the percentage of carbon, and which may often amount to a factor 2, is accompanied by a decrease in the deformability. When the welded structure is loaded, therefore, any deformation will have to be taken up to a greater degree by adjacent parts, so that an elongation may here occur beyond that which is permissible.

When it was finally understood that one of the most important causes of disappointments in the welding of St 52 was due to this phenomenon, the remedy was immediately obvious: the carbon used to obtain the high tensile strength of the steel should, at least partially, be replaced by other hardening elements whose action upon more rapid cooling did not, or not to that degree, increase. Manganese, chromium, silicon, copper, etc. may be considered. These elements, if present in too large quantities, may also have detrimental effects on the behaviour of the steel, so that it was necessary to establish maximum percentages, on the basis of experience, in which the hardening elements may be present in the steel St 52. These percentages are given in *table I*[1]).

Table I

Permissible percentages of different hardening elements in St 52 1), and percentages of the same elements in the material deposited from the Philips welding rod 52.

	St 52	Philips 55
carbon	0.20%	0.07%
manganese	1.50*)	0.90
silicon	0.50	0.30
copper	0.55	0.30
phosphorus	0.06	0.02
sulphur	0.06	0.02

*) 0.30% manganese may also be replaced by 0.40% chromium or 0.20% molybdenum.

These considerations, which until now have referred only to the parent metal, are also of importance for the welding rod itself. The iron of the welding rod with which we wish to weld St 52 must of course be provided with hardening components in order to obtain the desired high tensile strength (at least 52 kg/mm²). At the same time, however, in the most rapidly cooling parts of the weld the hardness must not be too great, so that the carbon content must be limited.

In the composition of the welding rod type 55 particular attention has been paid to the permissible limits for the proportions of the various hardening components. The values given in the second column of table I indicate that the deposited material of the rod remains far below the limits prescribed for St 52 itself[2]). Nevertheless the tensile strength is

[1]) Specifications of the German railways.

[2]) The relativity high content of copper is favourable for the corrosion resistance of the weld.

more than sufficient. In order to determine this a weld is made between two plates of St 52, and from this a flat test piece is made with the weld in the middle, see *fig. 1*. The tensile test gave a tensile strength of 55—57 kg/mm².

The tensile strength still depends to a certain degree upon the conditions during welding. In the test just mentioned the plates of St 52 steel were welded "horizontally", *i.e.* the bead lay horizontal. In "vertical" welding, beginning at the bottom and working upwards, the same test gave a somewhat higher value of the tensile strength, namely 59 kg/mm². If a weld is laid in a gutter of fairly thin sheet, and the weld metal thus obtained is turned to give round test pieces (fig. 1), lower values of the tensile strength are found. This is indeed quite understandable, since in this case the heat cannot flow away rapidly during the welding, so that the weld metal cools only slowly and is as it were annealed. With the Philips welding rod 55 a tensile strength of 48—52 kg/mm² was found with such rods and a yield value of 38—51 kg/mm². In fig. 1 it may also be seen that the weld metal possesses an unusually great deformability: the elongation at rupture (measured on a length 5 times the diameter) amounts to 35—40 per cent, while a constriction of 74—80 per cent occurs. This great elongation and constriction, which offer a guarantee that the welded structure can be considerably deformed due to strains without breaking (so-called plastic reserve of the construction) are due largely to the same measures as were taken to increase the notched bar toughness of the Philips rod 55. We shall discuss this in the following.

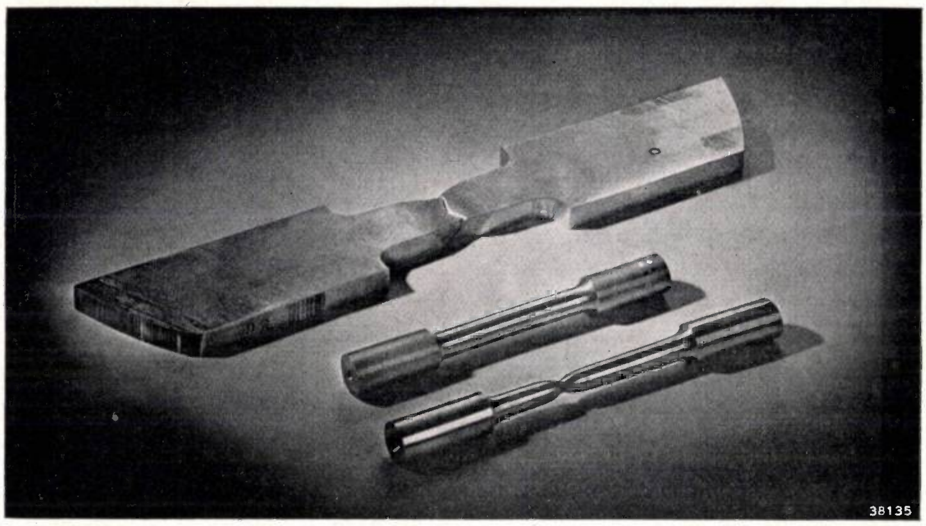


Fig. 1. Flat test piece for testing tensile strength made from a weld with the Philips welding rod 55 between two plates of St 52; the tensile strength of the steel was somewhat high, namely 60 kg/mm. Due to this the break occurred in the weld itself, as was desired, and not outside it. The round test pieces are made entirely of fused weld metal. The great elongation and pronounced constriction of the broken rods is striking.

The notched bar toughness of the weld

While the values of the tensile strength and yield value indicate only the maximum, or a point of the stress-strain diagram of special importance to the constructor, the value of the notched bar toughness gives a sort of integral over such a diagram, namely the total work which must be done by the deforming force in order to break the test rod into two pieces. This work is of especial importance for those cases where the deformations are concentrated at a single point of the piece of work, since then the rest of the piece does not as it were collaborate in taking up the work of deformation. Such a concentration (local increase of stress) occurs regularly at more or less sharp notches (this is the reason for the notching test), at transitions from this to thick cross section, etc.

In welds also such increases in stress will often occur, and it is therefore quite justifiable to require a high notched bar toughness for the deposited weld metal.

The notched bar toughness of a steel is found to be influenced to the highest degree by its content of nitrogen, as *fig. 2* shows [3]). The relation is of

[3]) Fig. 2 is borrowed from W. Bisehof and W. Püngel, Z. Westfälische Union Augustus 1939, p. 13. As to its general shape, the curve agrees witb observations of others, for instance D. Séférian, Diss. Paris, 1935, whose results have already been discussed by J. Sack in this periodical (Philips techn. Rev. 2, 129, 1937).

course not always the same, the further chemical composition also plays a part. Nevertheless the general conclusion is certainly correct that with increasing content of nitrogen the notched bar

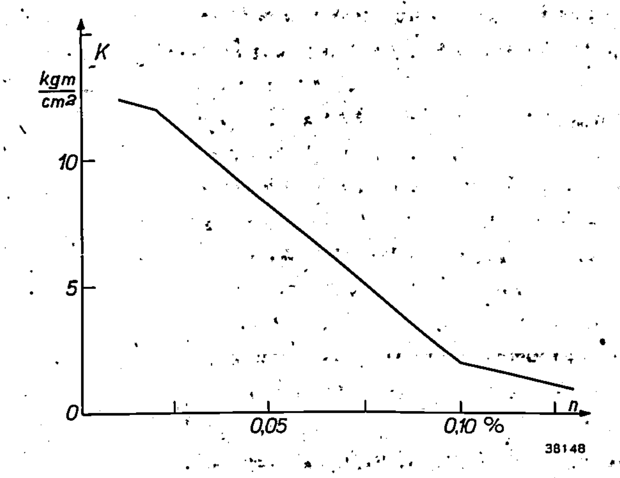


Fig. 2. Variation of the notched bar toughness k as a function of the nitrogen content n in per cent by weight [3]).

toughness decreases. Good qualities of steel thus contain only little nitrogen. In welding, however, there is a danger that nitrogen in atomic form will be taken up from the air (at the high temperature of the welding arc the air is for the most part dissociated into atomic oxygen and nitrogen). Combatting this is one of the most important functions of the coating of the welding rod. In the first place this coating forms a more or less sealing layer of slag on the fused iron. Furthermore the coating can be so composed that upon heating it develops large amounts of gases, harmless in themselves, which form a protecting atmosphere around the fused material. Especially organic components, such as starch, sawdust etc. are used for this purpose. In the case of the rod type 55 a different method has been chosen to reinforce the nitrogen resisting action of the coating, namely the addition of substances which are able to form a chemical compound with the nitrogen (and also with oxygen). Thanks to these measures the percentage of nitrogen in the deposited weld metal could be reduced to the very low value of 0.01 per cent. For the sake of comparison the nitrogen content of several other common kinds of welding rods is given in *table II*. At the same time it may be seen that the expected strong increase in the notched bar toughness runs parallel to the decrease in the nitrogen content; in the case of the Philips rod 55 the former amounts to 17 to 20 kgm/cm² on a cross section to be broken of 5×10 mm², and 73 to 26 kgm/cm² on a cross section of 2×10 mm².

Table II

Nitrogen content and notched bar toughness of the deposited material of different kinds of welding rods.

Type of welding rod	Nitrogen content in %	Notched bar toughness (5×10) in kgm/cm²
Uncoated rod	0.15	1
Medium thickly coated rod	0.07	4
Thickly coated normal quality rod	0.035	10
Medium thickly coated rod with partly organic coating	0.0125	11
Thickly coated rod Philips 55	0.010	18

The values of the notched bar toughness mentioned are measured on impact test pieces which were prepared as shown in *fig. 3* from the welded joint. The notch is in this case so far away from the parent metal that the latter has no effect on the result. The good reproducibility of the impact test is striking. *fig. 4* shows a series of 12 rods with a cross section of 5×10 mm² which have been subjected to the impact test. Their notched bar toughness or impact value was found not to vary more than from 17.2 to 10.5 kgm/cm².

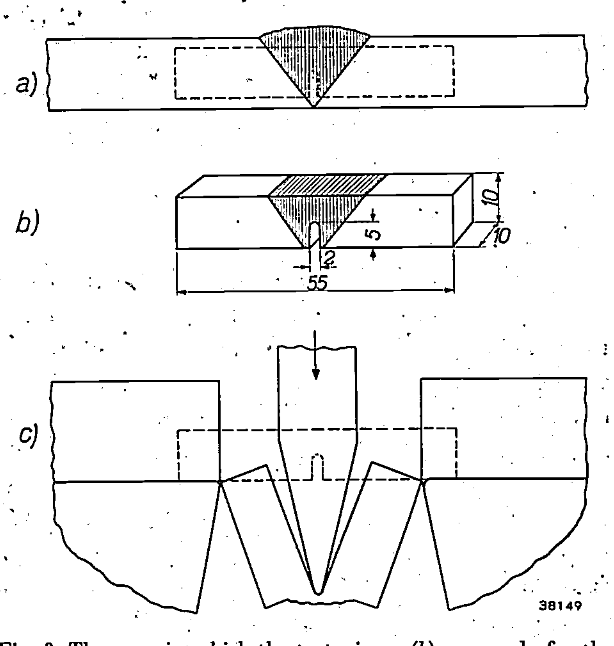


Fig. 3. The way in which the test pieces (b) are made for the impact test (c) from a weld (a). The deposited weld metal is shaded.

In giving these values it must also be mentioned that they are obtained at room temperature (about 20° C). In contrast to all other properties of materials of interest to the constructor, the notched bar toughness is very closely dependent upon the temperature, and this dependence is similar for all kinds of steel: there is a pronounced minimum at 400—500 °C and a rapid fall at temperatures below the zero point. The significance of this sharp fall

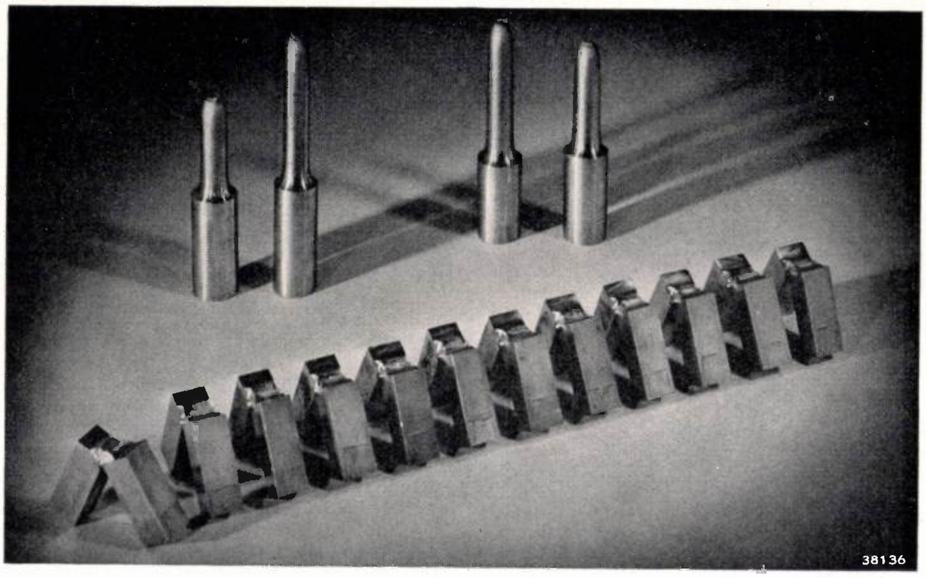


Fig. 4. Series of 12 test pieces made from welds with the Philips welding rod 55 to which the impact test has been applied. The results are seen to be satisfactorily reproducible. The pieces are not broken through, but bent down between the supporting blocks by the impact. In order therefore to obtain the correct impact toughness a small (unknown) amount of work which would be necessary to separate entirely the two halves of the test piece must still be added.

in notched bar toughness at low temperatures (the so-called cold brittleness) was formerly often not sufficiently realized. It has been one of the causes of various accidents in which welded bridges have developed cracks during very cold weather [4]). If one compares the variation of the notched bar toughness for the material deposited from the welding rod 55 with the curve for a normal good quality welding rod, *fig. 5*, it may be seen that the much higher value of this toughness has as it were shifted the curve as a whole to higher values. While in the case of the ordinary welding rod the notched bar toughness decreases from 11 kgm/cm² at 20° to 6 kgm/cm² at —50 °C, in the case of the Philips rod 55 the notched bar toughness at —50° C still has a value of about 14 kgm/cm². Only at still lower temperatures, which, however, do not occur in inhabited parts of the world, would the notched bar toughness fall to dangerously low values.

In other respects also besides temperature, the notched bar toughness is a particularly "sensitive" property of materials. In the heat treatment to which welded constructions are sometimes subjected in order to remove stresses occurring as a result of the welding, the notched bar toughness of some kinds of welding rods is found to decrease enormously, even to $^1/_{10}$ the original value, while other properties, such as tensile strength and elongation, exhibit only relatively small changes which are quite in line with the expectations. Such a great decrease in the notched bar toughness must be ascribed mainly to a too large percentage of nitrogen in the rapidly cooled weld metal [5]). The nitrogen

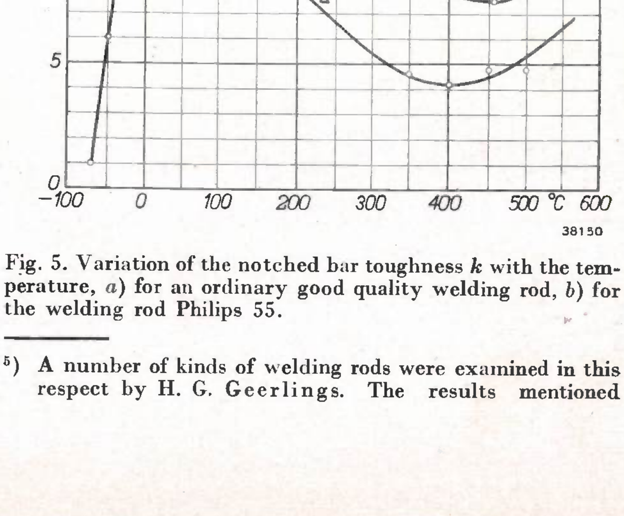


Fig. 5. Variation of the notched bar toughness *k* with the temperature, *a*) for an ordinary good quality welding rod, *b*) for the welding rod Philips 55.

[4]) This was the case of the bridge of the Reichsautobahn at Rudersdorf and the two bridges at Herenthals and Kaulille over the Albert Canal.

[5]) A number of kinds of welding rods were examined in this respect by H. G. Geerlings. The results mentioned

is separated out on the crystal boundaries in the form of Fe_4N upon reheating, and this produced a more brittle material. In agreement with this explanation is the fact that this effect is very small when the welding is done with the Philips rod 55, where the weld metal contains very little nitrogen. Even after the so-called normalizing annealing (at 920° C) the original high value of the notched bar toughness is still measured.

The separation of nitrogen mentioned above also takes place at room temperature, although extremely slowly. The so-called ageing of the weld metal must be ascribed to this phenomenon (among other factors). Since experiments in this direction would have to extend over a very long period, an artificial ageing test is usually applied for judging materials in this respect. By stretching the material (10 per cent) and then annealing it ($^1/_2$ hour at 250 °C) a kind of accelerated ageing is obtained. In the case of the weld metal of the rod 55 this was found to cause only a slight decrease in the notched bar toughness, namely from 17.7 to 13.5 kgm/cm², while for a normal good quality welding rod a decrease from 10.4 to 4.5 kgm/cm² was found.

In order to complete the picture of the properties of the Philips welding rod 55, it may be mentioned that the fatigue bending strength (load necessary for breakage after 5×10^6 alternations) amounts to 26—28 kg/mm², a value in the neighbourhood of that for St 52; furthermore that the welds can easily be forged and that the bending test also gives good results. Several rods which were subjected to the latter test are shown in *fig. 6*, the text beneath the figure gives further details.

Fig. 6. A commonly applied method of testing welded joints is the bending test. A test piece with the weld at its middle is bent over a mandrel of a given diameter until cracks occur. The test pieces here shown are 10 mm. thick and made of St 52 welded in a V-weld with the Philips rod 55 and bent over a mandrel of 30 mm diameter. With a bending angle of 180° the weld is still intact, even in the test piece in the foreground which was flattened after being bent.

here were published by him in Laschsymposium 1940-1941, p. 13. It must be mentioned that the oxygen also plays a part in the decrease in notched bar toughness, and that there is almost always 2 to 3 times as much oxygen as nitrogen in the material fused. Much less, however, is known about the part played by the oxygen than that of the nitrogen.

The welding of "difficult" types of steel

While in the above it was assumed that the deposited weld metal must be the equal in properties of the parent metal, cases also exist in which this requirement is unnecessary, and where nevertheless difficulties are experienced in welding. If for example one tries to weld special kinds of steel such as StC 45 (a very hard steel with a carbon content of 0.45 per cent), StC 60, etc. with an ordinary good quality welding rod, aftercooling cracks often appear in the weld. Upon closer consideration this is not strange. Aside from the great increase in hardness in the transition zone, which is caused by the high carbon content of the basic material, the latter of itself is already so hard and little deformable that even upon shrinking during cooling very high requirements are made of the deformability of the weld metal. Thanks to the great toughness and the considerable plastic reserve of the material deposited by the Philips rod 55, it is also possible to weld the difficult, very hard steels with this rod. This is demonstrated by *fig. 7*, while in *figs. 8* and *9* a practical example is reproduced where use has been made of the possibility mentioned. In the cast steel head of a heavy press (300 tons pressing force) a 50 mm thick reinforcing plate of St 37 had to be welded. In spite of the high carbon content of the cast steel (0.60 per cent) and the large dimensions of the parts to be welded, a weld which was free of cracks could be obtained. In order to limit the speed of cooling the head of the press was heated to 125 °C before welding.

Another kind of difficulty which may occur in welding is the formation of gas bubbles in the transition zone, usually to be ascribed to impurities in the layer of the parent metal which becomes mixed with the fused weld metal. If these gases are liberated at an unfavourable moment, namely during the solidification of the material, they can under certain circumstances no longer escape, or the fused material can no longer fill up the cavities, and a porous weld is obtained with cavities and craters. It is unnecessary to state that this is very detrimental to the mechanical properties of the

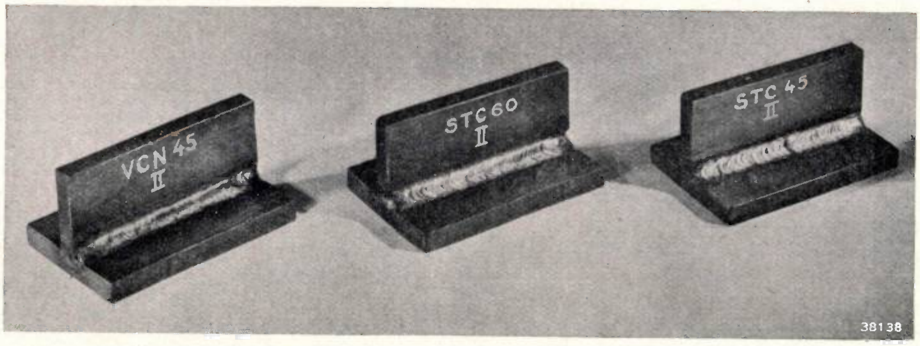


Fig. 7. Fillet welds on three different kinds of very hard steel. The welds are made with the Philips rod 55 according to a standardized method of the German railways for testing for freedom from cracks. Except for a single very small crater crack no cracks were found in any of the three welds reproduced here.

welded joint, especially because cracks may develop from the cavities.

It has been found by experience that too large a content of sulphur in the parent metal must often be held responsible for this formation of gases and the resulting porosity [6]). It is known from metallurgy that in the Siemens-Martin process a removal of sulphur from the iron is obtained by means of lime in a reducing environment. The sulphur is then for the most part bound by the calcium; for which the following reaction may be accepted:

$$FeS + CaO + C \rightarrow Fe + CaS + CO.$$

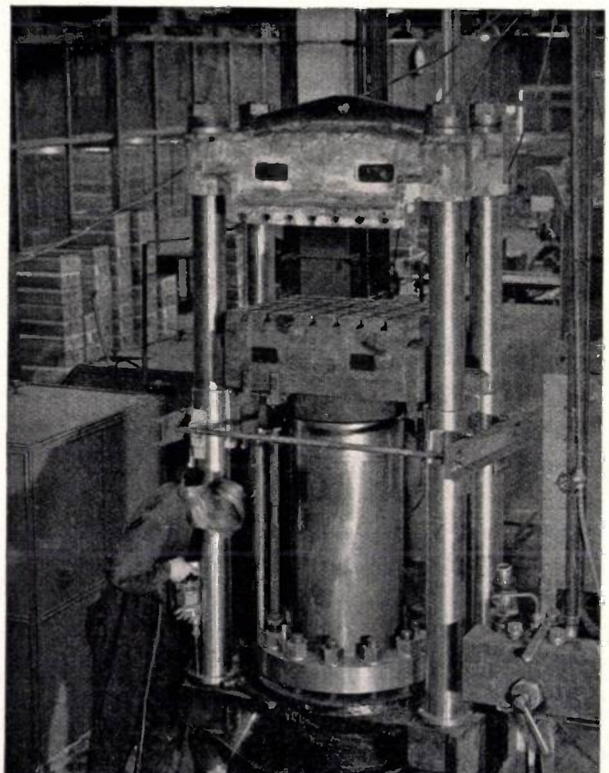


Fig. 8. Press with 300 tons pressing force during assembly. In the head of the press a reinforcing plate 50 mm thick has been welded with the Philips rod 55. The head is of cast steel (0.60 per cent carbon), the reinforcing plate of St 37.

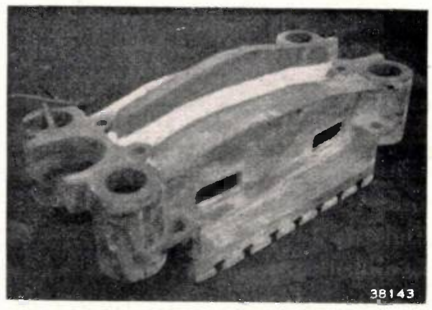


Fig. 9. Head of the press of fig. 8 with welded reinforcing plate.

The calcium sulphide formed, in contrast to the iron sulphide, is entirely taken up in the slag, and the sulphur content of the steel is thus decreased. The coating of the Philips rod 55 is also composed on a lime basis, so that in the fused material, which is in intimate contact with the slag, a binding of any sulphur present in the basic material can take place. This was well illustrated in a test in which a steel poor in sulphur and one rich in sulphur (0.03 and 0.23 per cent, respectively) were welded with the Philips rod 55, while afterwards the slag was chemically analysed. In the second case the slag was found to contain 2 to 3 times as much suphur

[6]) An important example from practice may be found in L. Reeve, Trans. Inst. Welding 3, 11, 1940. A report is here given on the case of the bridge near Hasselt (Belgium) which collapsed several years ago. Photographs of preparations made from the welded points of this bridge show gas cavities at spots immediately adjacent to the parent metal which contained 0.095 and 0.063 per cent of sulphur, respectively.

as in the first. The fact that the Philips welding rod 55 is indeed very insensitive to sulphur is clearly shown by *fig. 10* where fillet welds are shown which were laid on a steel with much sulphur (0.09 per cent) with the Philips rod 55 and with an ordinary good welding rod, which was, however, sensitive to sulphur. The difference is even more pronounced when one attempts to weld free-cutting steel. This steel contains very much sulphur in the form of FeS, to which its easy machinability is due, and which makes it suitable for machining in automatic lathes, etc. *Fig. 11* shows a fillet weld in free-cutting steel with 0.23 per cent of sulphur (and in addition 0.08 per cent of phosphorus), the right-hand half of which is welded with an ordinary good quality welding rod and the left-hand half with the Philips 55. The latter half is absolutely tight, and, as shown by stretching and bending test, mechanically quite satisfactory, while the weld obtained with an ordinary rod is quite porous and mechanically useless. It is therefore possible to state without exaggeration that with the Philips welding rod 55 types of steel can be welded which have been characterized as unweldable up to the present.

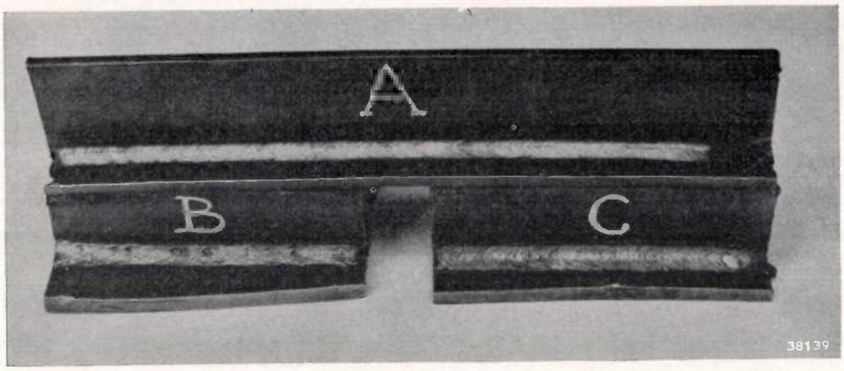


Fig. 11. Fillet weld in free-cutting steel (0.23 per cent sulphur, 0.08 per cent phosphorus). The right-hand half is welded with an ordinary good quality welding rod (quite porous weld), the left-hand half with the Philips rod 55 (tight weld).

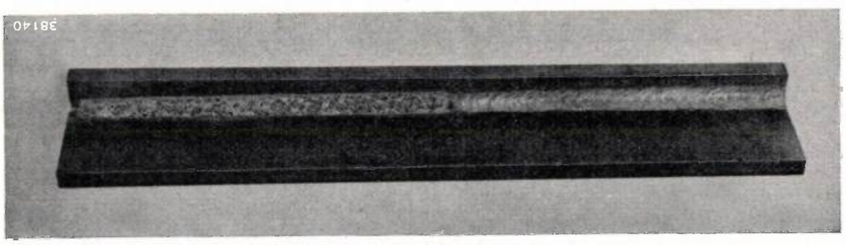


Fig. 10. *A* fillet weld made with an ordinary good quality welding rod on St 37 with normal sulphur content (0.03 per cent). *B* fillet weld with the same rod on steel rich in sulphur (0.09 per cent); holes may clearly be seen in the weld which has a "burned" appearance (so-called pock-marked). *C* fillet weld on the same steel rich in sulphur made with the Philips rod 55; the weld is quite sound.

INCANDESCENT FILAMENT LAMPS FOR SERIES-CONNECTION

by N. A. HALBERTSMA and J. A. M. van LIEMPT. 621.316.062.2 : 621.326

Although it is at present customary to connect electrical apparatus in multiple to the supply main, series connection of lamps may offer advantages, under certain circumstances. In order to avoid extinction of the whole installation when one lamp becomes defective special measures are taken which are described in this article.

Introduction

The advantages of multiple connection of the lamps in a system of illumination over the series connection which was formerly often used, consist chiefly in the fact that the lamps are independent of each other. This is usually considered to be of such great importance that in designing illumination systems no other possibility is as a rule considered. By way of exception, however, connection in series is also sometimes used, in spite of the fact that the possibility of switching on any desired number of lamps and the free choice of the size of the lamps must be sacrificed.

The use of series-connection is justified when:

1) the available voltage is too high for the lamps to be used, and
2) the simplification obtained is important.

All cases in which connection in series is used, as the lighting of trams and trains, illumination and Christmas tree lighting and, to a limited extent, street and highway lighting, can always be classified in one of these two groups.

The available voltage is too high for the lamps to be used.

In the case of A.C. installations the voltage can be lowered to any desired value in a simple manner with the help of a transformer. In the case of D.C. installations, however, the voltage is fixed. It is for this reason that the connection of electric lamps in series is usually used for the lighting of trams and trains. The D.C. voltage on the trolley wire is usually 500 to 600 volts; the highest voltage for which electric lamps are made is 260 volts. As a rule with a voltage of 600 volts 5 lamps of 120 volts are connected in series; sometimes a larger number of lamps of a lower voltage are preferred. If the traction voltage is 1 500 volts and higher it is advisable not to introduce this voltage in parts of the coaches, where is would be accessible to the public. It is better to use a converter to supply current to a distribution system for lighting and other purposes, and then ordinary multiple connection can be used.

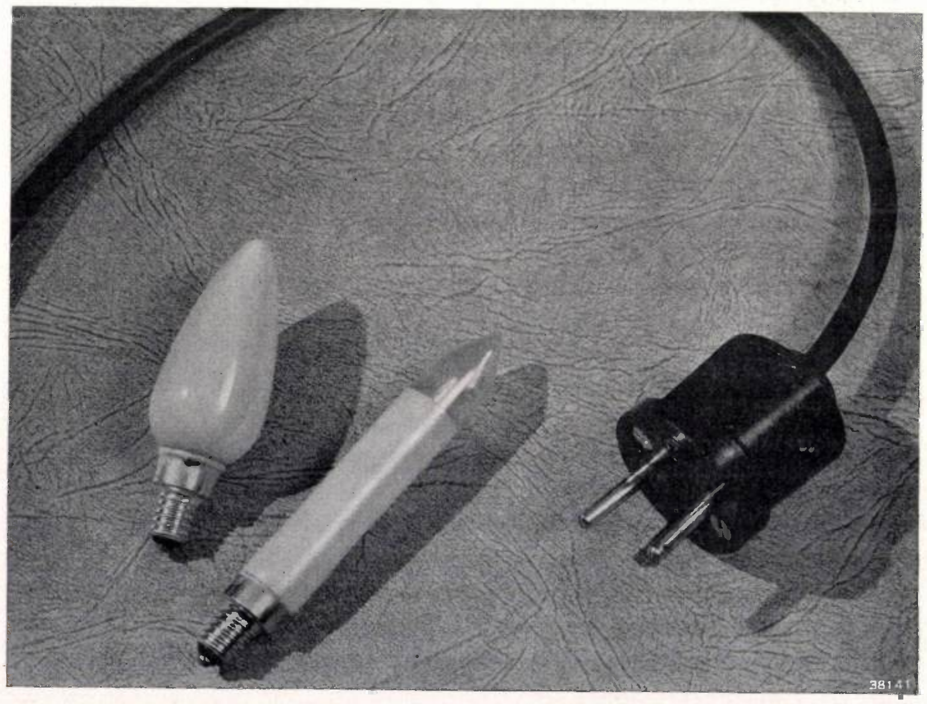


Fig. 1. Illumination and Christmas tree lamps, both for 14 volts, 3 c.p.

The ordinary mains voltages of 230 volts and 115 volts may also in certain cases be too high for the lamps which must be used. The smallest ordinary electric lamp for 230 volts and 115 volts has a light flux of 15 dlm, and a consumption of about 15 watts. Lamps with smaller wattage, such as those for illuminations and Christmas trees, must be made for a lower voltage because of the limit imposed by the thickness of the wire. A voltage of 14 volts with a light intensity of 3 c.p. is the most common. With a consumption of 3.75 watts these lamps have a current of 0.27 A. The filament is short and stout, the lamp which is small (*fig. 1*) may be provided with a small-sized screw cap. The lamp holder is thus also small, which is important since the lamps are used for decoration. On 220 volt mains 16 such lamps can be connected in series, and on mains of lower voltage a correspondingly smaller number.

The simplification of the installation

In connection in series there is only one conductor which passes from socket to socket, while in multiple connection two wires must branch off to every lamp socket. This makes the installation so much simpler that in certain cases series-connection will be preferred. The illumination of a Christmas tree is effected more rapidly with a single thin flex than with the thicker double flex which must be used for multiple connection.

The simplification of the installation is also important in the installation of street lighting systems with series-lamps as is often found in the United States. For new installations, however, in the United States preference is usually given to connection in parallel. In Europe the connection in series for street lighting has only been used on a large scale in Italy.

When a very long traject must be illuminated with light sources of small power and considerable spacing, the connection in series may offer advantages with respect to the cost of installation. Along both sides of the Amsterdam - IJmuiden Canal for 15 miles there is a series installation with lamps of 65 watts spaced at 900 ft.

Disadvantages of series-connection in practice

In addition to the objection already mentioned that with connection series it is impossible to change at will the number of lamps burning, there is the further disadvantage that if one of the lamps becomes defective the whole series is extinguished. In a tramcar, for example, this would be very inconvenient, since one is then compelled to hunt for the burnt out lamp in the dark in order to replace it. The lamp which has failed is, however, indistinguishable from the others, and the hunt for the defective lamp has long been an unpleasant aspect of the use of series-connection of electric lamps.

For this reason devices have been applied, particularly for street lighting, for the automatic short-circuiting of a defective lamp. Two aims are hereby achieved, namely

1) that all the lamps of the series except the defective one continue to burn, and
2) that the defective lamp can easily be found since it is the only one extinguished.

In order to maintain the amperage entirely unchanged, the burnt-out lamp should actually be replaced by an equivalent resistance. If, however, the lamp is short-circuited, the resistance of the whole series will decrease and the current will increase slightly. The smaller the number of lamps in series, however, the greater the overloading of the remaining lamps. In tramway cars where for instance 5 lamps of 120 volts burn in series, the voltage per lamp would rise by 25 per cent, *i.e.* to 150 volts. It is therefore necessary to take measures to bring the current quickly back to its normal value in order to avoid the failure of more lamps due to the excess voltage. If such measures are impossible, it is advisable to use lamps with a lower voltage. Voltages of 24, 30, 40 and 50 volts may be used.

Automatic short-circuiting arrangements for burnt out series lamps

For the automatic short-circuiting of a lamp in a series very different devices have been proposed. Use has for instance been made of the interruption of the current to short-circuit the lamp by means of a relay, or to replace it by an equivalent resistance. The interruption of the light may in this case be of such a short duration that scarcely more than a flicker is observed.

The rise in voltage at the terminals of the defective lamp may also be used to bring about short circuiting, due to the fact that a very thin layer of insulation material held between two metal contacts breaks down. In the United States such devices are used as a rule in series street lighting systems. The flash-over device consists of two aluminium discs one of which is provided with a thin insulating layer which breaks down at about 400 volts. The disc is clamped between two contact springs fastened to the lampholder, which contact springs serve at the same time to make the connection between the lampholder and the terminals.

When a lamp burns out the lamp with the holder can be removed from the fitting and the lamp and flash-over device both renewed without any risk of touching parts under tension.

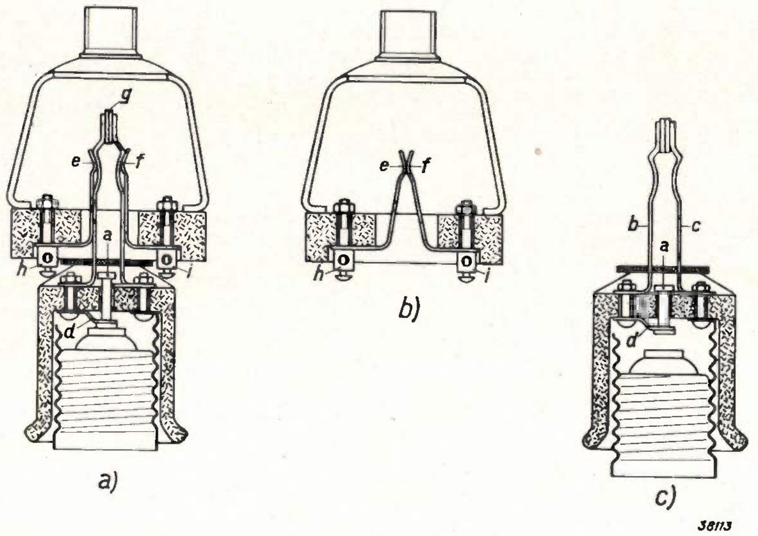


Fig. 2. Lamp-holder with short-circuiting device for street lighting with series-lamps, *a*) ready for use, *b*) lamp-holder with lamp and voltage fuse removed and *c*) lamp-holder with voltage fuse and lamp. *a* contact, *b* and *c* disc holder, *d* flexible strip, *e* and *f* contact springs, *g* flash-over disc, *h* and *i* terminals.

Screwing out a lamp which forms part of a series installation is dangerous, not only because of the risk of touching a part under tension, but also for another reason. If the current continues to flow, especially in the case of direct current, an arc will occur between the centre contact of the holder and the centre contact of the screw cap, which will melt the metal parts mentioned. The lamp holder shown in *fig. 2* has another short-circuit contact for preventing this, which is connected to the flexible middle contact and comes into action as soon as the lamp has been screwed out of the lampholder a fraction of an inch, and before contact with the centre contact of the cap is broken.

In analogy with the fuses which come into action when a definite current value is exceeded, the flash-over devices may be called "voltage fuses", since they come into action as soon as a definite voltage is reached, which is higher than the rated voltage of the lamp but lower than the mains voltage.

The voltage fuses, which may be incorporated in the lamp itself or connected separately to the lamp-holder, must satisfy very divergent requirements. At the burning voltage of a single series-lamp they must have an infinite resistance, but at the total voltage applied to the series they must break down and restore the circuit. In street lighting with series-lamps this is not difficult to achieve, due to the high voltages occurring. There are numerous insulating materials which will break down when exposed in a very thin layer to voltages of 3000-10 000 volts, while at several hundred volts they exhibit practically no conductivity.

In tramcar lighting the breakdown voltage must lie below 400 to 500 volts and above the rated voltage of the lamp, and when, to mention another example, a series of lamps for illumination of 7 or 14 volts is connected to a main of 115 volts, the voltage fuse must operate butween 20 and 100 volts. Very thin insulating layers are necessary for this, and they should have little spread in the breakdown voltage. It is indeed possible to produce such fuses.

For the operating of the voltage fuses it is a matter of some importance whether the lamps in series are vacuum lamps or gas-filled lamps. When the filament fails of a vacuum lamp which is not provided with a voltage fuse no electric discharge can take place, due to the absence of gas molecules in sufficient quantities, the vacuum acting

Fig. 3. Photograph of a series lamp which has been destroyed by an arc discharge.

as an insulator and the current is simply interrupted. The case is different with a series gas-filled lamp without a voltage fuse. In this case, when the filament of the burning lamp fails, the current need not immediately be interrupted. An arc may occur at the break of the spiral which heats the ends to such a high temperature that they begin to melt. The other lamps of the series now serve as a resistance, stabilizing this arc discharge. It sometimes occurs that the spiral is entirely fused away and that the arc strikes over to the leading-in wires of the lamp. The latter then also fuse, so that, unless the lamp has already been destroyed by the heat developed, the arc finally arrives in the neighbourhood of the foot (*fig. 3*) and not only destroys the glass with the leading-in wires, but may also pass over to the lamp cap and even to the metal parts of the lamp holder. Such an arc may, especially with direct current and under conditions favourable to its maintenance, cause considerable damage.

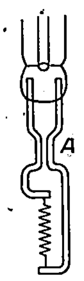


Fig. 4. In order to promote the rapid occurrence of short circuiting by the arc the leads of this series lamp are brought close together at *A*.

This process can, as stated, only take place in the absence of a voltage fuse or when the latter does not work, of which there is a certain chance since the voltage of the arc may remain below the breakdown-voltage of the voltage fuse. The solution of this difficulty might be sought in raising the voltage of the arc when it occurs by placing the leading-in wires as far apart as possible. This would, however, result in difficulties in connection with the introduction of the filament system through the narrow neck of the bulb; the solution has therefore been sought in just the opposite direction, namely by bringing the wires very close together. If an arc is produced the leads will fuse off until at *A* in *fig. 4* the drops of molten metal on the ends of the nickel wires short-circuit the lamp and thus prevent both a break of the circuit and damage to the lampholder.

If the lamp has very short leading-in wires which are difficult to bend close to each other, they are provided with separate wires which serve as "arc catcher" (*fig. 5*). Even when these measures have been taken a voltage fuse is not superfluous. If for example a spiral breaks (by shock, for instance) while the lamp is not burning, no arc will occur when the series of lamps is switched on. In this case the voltage fuse must break down and short-circuit the defective lamp before the series of lamps actually lights.

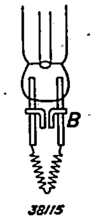


Fig. 5. Series-lamp with separate arc-catcher *B*.

The type of voltage fuse varies according to the voltage at which it must act. The piece of thin paper formerly employed has been universally replaced by an insulating material which is deposited on metal. Metal oxides which are good insulators are preferred. In the oxidation process it is possible to control the thickness of the oxide film. A film of aluminium oxide is usually used which with the dimensions used for a breakdown voltage of about 400 volts has a resistance of 10^8 ohms, and may thus be considered as an insulator. After breakdown this resistance falls practically to zero. For illumination and Christmas tree lamps the breakdown voltage must be so low (below 100 volts) that aluminium oxide cannot be used. The film would have to be so thin that it is not to be realized in practice. Therefore a sodium tungstate is used in these cases, the so-called blue tungsten bronze, which still has a low breakdown voltage at a greater thickness of the layer.

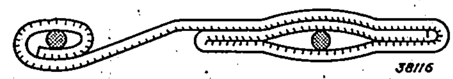


Fig. 6. Metal-strip covered with aluminium oxide which is bent around the leading-in wires in such a way that only one layer needs to break down to cause a conducting connection between the wires.

In the case of vacuum lamps the voltage fuse is mounted between the screw cap and the glass of the bulb. It cannot be placed in the bulb itself because the oxide films used contain gases which would spoil the vacuum of the lamp. In gas-filled lamps the gas freed from the oxide is of no importance. The strip with aluminium oxide can be bent around the lead-in wires (*fig. 6*), if desired combined with a nickel strip which serves as arc catcher (*fig. 7*). This voltage fuse shown in fig. 7, may be seen in *fig. 8* mounted inside the bulb of a gas-filled series lamp for tramcar lighting. The voltage fuse of fig. 6 may be seen in *fig. 9* mounted in the screw cap of a vacuum series lamp which is shown in its true dimensions.

In the case of illumination and Christmas tree lamps the voltage fuse, in spite of the small space available, can be placed in the tiny screw cap due to its simple construction. An example is shown in *fig. 10*.

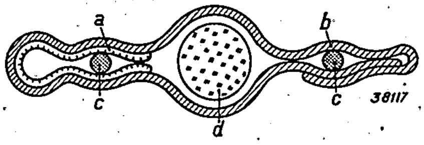


Fig. 7. Combination of an aluminium strip (a), provided with an insulating layer of aluminium oxide with a nickel strip (b), which serves as arc catcher. c leading-in wires, d glass rod.

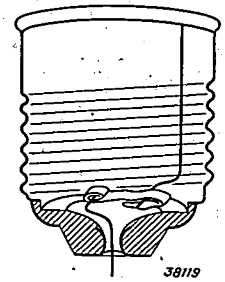


Fig. 9. In vacuum series-lamps the voltage fuse shown in fig. 6 is mounted into the screw-cap. The sketch is full size and shows the small dimensions of the voltage fuse.

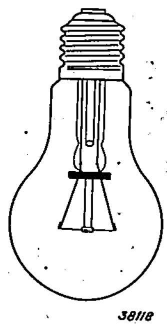


Fig. 8. Series-lamp (gas-filled) for tramcar lighting for 40 volts, 40 watts (1 amp.), in which the voltage fuse shown in fig. 7 has been mounted.

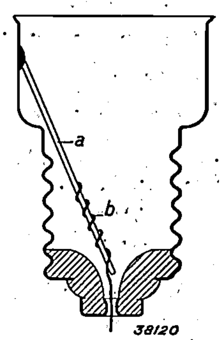


Fig. 10. Miniature screw-cap for small illumination lamps, twice natural size, in which an extremely simple voltage fuse has been mounted. a tungsten wire covered with tungsten bronze, around which a metal wire b is wound.

NON-LINEAR DISTORTION OF SOUND FILM WITH OBLIQUE LIGHT SLIT

by J. F. SCHOUTEN.

778.534.45 : 621.396.813

When a sound film is recorded or scanned with a light slit which is not exactly perpendicular to the direction of length of the film, a non-linear distortion of the sound occurs. An exact calculation of this distortion was considered impossible until now. In this article it is shown that the calculation is indeed possible, and even relatively simple, when the light diffraction at the sound film is used for assistance. The diffraction spectrum consists of a two-dimensional pattern whose light distribution along a given axis (parallel to the direction of the film) provides the Fourier spectrum of the sound obtained upon reproduction. With an "obliquely recorded" film the diffraction pattern is also found to be obliquely distorted, and with this very simple transformation the new Fourier spectrum can immediately be derived, when the diffraction pattern of the "vertically recorded" film is completely known. In this way the distortion with oblique slit is calculated and discussed for the cases where one is concerned with only one frequency, and where one is concerned with two different frequencies in the sound track.

The problem of the oblique slit in sound film technology

If upon reproducing sound film the narrow light slit with which the film is scanned is not perpendicular to the direction of length of the film (*fig. 1a*), then when the sound track on the film is sinusoidal the amount of light transmitted does not vary truely sinusoidally with the time (fig. 1*c*). A non-linear distortion occurs which is manifested acoustically in that the sound contains higher harmonics in addition to the fundamental tone.

A similar phenomenon will occur when in recording the film the light slit (in the photographic system) or the cutter (in the Philips-Miller system, where the sound track is cut in the film [1])) is not perpendicular to the direction of length of the film. A sound track results which as it were slants (fig. 1*b*), and which upon reproduction with a straight slit leads in the same way to non-linear distortion.

The fact that upon oblique scanning non-linear distortion must occur was already pointed out in 1881 by König [2]) in the discussion of his "wave syren", a forerunner of the sound film. The problem only became actual in sound film technology. Frieser and Pistor [3]) found in their attempts at an exact calculation that even with a single frequency they encountered unsolvable equations, and they confined themselves to working out several cases graphically. Podliasky [2]) gave an approximate calculation, also for the case of a single frequency, for the relative intensity of the second harmonic which occurs at very small angles of the slit. The problem of the general calculation of the distortion was, however, considered insoluble until now.

We shall show how with the help of the previously described light diffraction by sound film [3]) an exact and, moreover, very clear solution of this problem can be obtained in a relatively simple way.

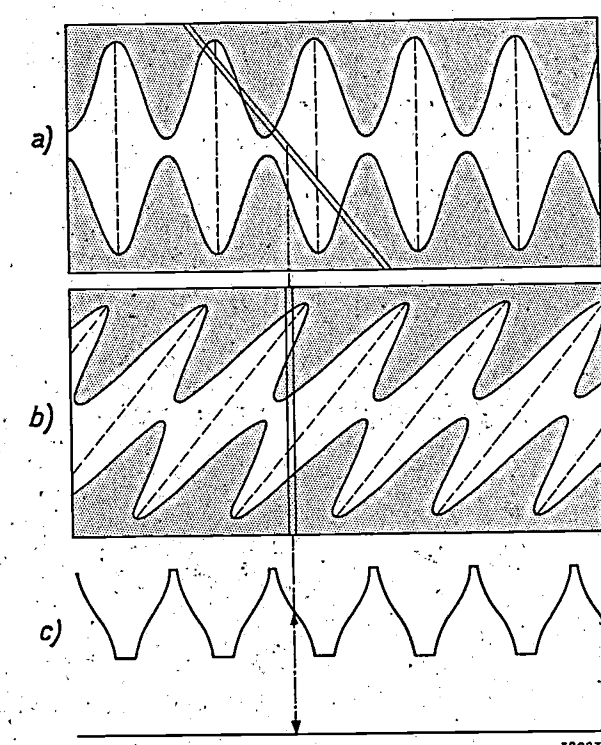


Fig. 1. If a piece of film is scanned with a light slit which is oblique with respect to the direction of length of the film (*a*) or if the sound track is recorded with an oblique slit (*b*), in both cases non-linear distortion of the sound results: the amount of light transmitted with a sinusoidal sound track is no longer sinusoidal but as shown in (*c*).

[1]) R. Vermeulen, Philips techn. Rev. **1**, 107, 1936. The strips of film reproduced in this article were all recorded by this system. For the sake of simplicity, however, we shall always speak of a light slit.

[2]) R. König, Wied. Ann. Phys. **12**, 335, 1881.
H. Frieser and W. Pistor, Z. techn. Phys. **12**, 116, 1931.
I. Podliasky, Annales P.T.T. **24**, 1, 1935.

[3]) J. F. Schouten, Philips techn. Rev. **3**, 298, 1938; **4**, 290, 1939.

Before passing on to the actual calculation we shall indicate the relation between light diffraction by sound film and non-linear distortion in a general way.

General consideration of the diffraction pattern of obliquely recorded sound film

We have previously described [3]) how a strip of sound film can be used to produce optical interference phenomena. A diffraction pattern is obtained which is also important acoustically. The very general proposition can be deduced *that the light distribution on a given line in this diffraction pattern* (namely the horizontal axis through the middle of the pattern, indicated below) *gives the exact Fourier analysis of the sound recorded on the film* [4]).

In *fig. 2* a strip of film is shown on which a tone of 400 c/s has been recorded, with the diffraction pattern obtained from it. Along the horizontal axis only three lines are encountered, a central line (zero order) and lines to the left and right of it (first order), which means that the strip of film contains only the frequencies zero — the average width of the film track may be considered as such — and 400.

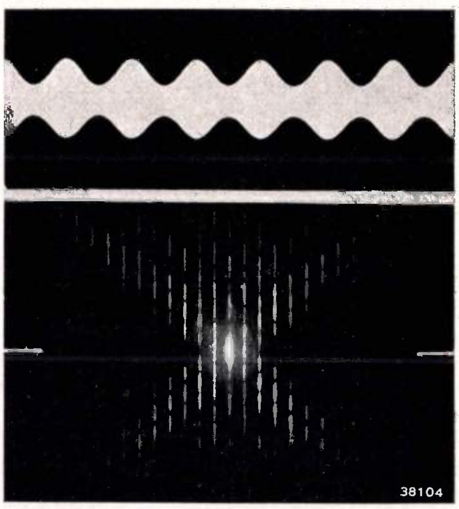


Fig. 2. Light diffraction by sound film. Above: sound track (tenfold enlargement) of a tone of 400 c/s recorded with vertical slit. Below: the diffraction pattern obtained. On the horizontal axis only the lines of the zero and first orders are visible.

The pattern outside the axis here has no acoustic significance. While the pattern actually also contains lines of higher orders (*i.e.* at greater distances from the centre), since these lines do not occur on the horizontal axis, the sound recorded on the film does not contain the corresponding frequencies of 800, 1 200, etc.

Let us now consider the diffraction pattern of a strip of film upon which the same frequency is recorded obliquely (*fig. 3*), with a (very exaggerated) angle of 15 degrees. The pattern again

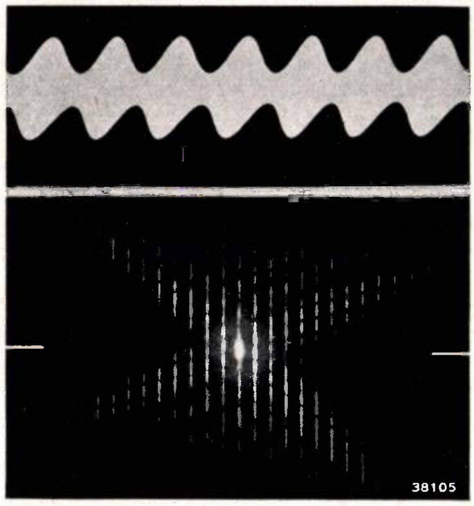


Fig. 3. The same as fig. 2, but for recording with an oblique slit. The diffraction pattern is identical with that of fig. 2 except for a displacement of the lines in a vertical direction.

contains the same lines of higher order, each line with the old intensity distribution in the vertical direction [5]), but shifted in that direction over a distance proportional to the number of the order.

This results in the fact that the second, third and fourth orders are shifted on the horizontal axis with appreciable intensity, and therefore upon reproduction will occur in the sound as second, third and fourth harmonics.

The optical spectrum of a vertically recorded sound track thus gives us complete insight into the sound spectrum which will be obtained upon oblique recording or reproduction. *The higher orders are as it were ready to slide across the horizontal axis upon oblique recording and to become observable in the sound as higher harmonics.*

In principle quite the same behaviour is found with more complicated vibration forms (figs. 17 and 18). Here also the general structure of the pattern remains intact, the lines, however, are shifted along each other and may thereby cause an appreciable intensity on the horizontal axis.

The calculation of the non-linear distortion caused by an oblique position of the light slit thus falls into two steps: the calculation of the normal diffraction pattern and the calculation of the shifts

[4]) J. F. Schouten, Physica 7, 101, 1940.

[5]) Except for a slight difference which may be ascribed to the slightly wider zero track and the slightly greater amplitude of the sound on this film.

which this pattern undergoes due to the oblique position of the light slit [6]).

Calculation of diffraction pattern

The diffraction pattern of the perforated grating

Consider a parallel beam of light which is incident perpendicularly on the surface of the film. If there is a small hole in the film the light is diffracted uniformly in all directions by this hole when we confine ourselves to small angles of diffraction.

The coordinates of the rectangular hole are u and v, its width and height du and dv, respectively, and the angles of diffraction in horizontal and vertical direction α and β, respectively. We consider first the diffraction of the rays in the horizontal direction and draw the projection on the horizontal plane through the u axis, see *fig. 4*. In this drawing a ray through the point u is given which is diffracted through a horizontal angle α and a similar ray which would pass through the point zero.

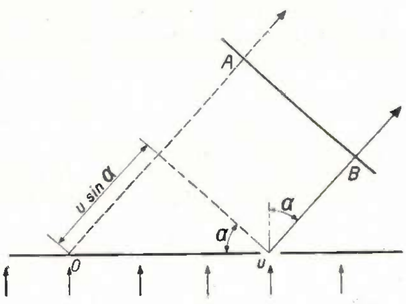


Fig. 4. Light diffraction at a hole (projection in the horizontal plane).

On an arbitrary line AB perpendicular to the path of the rays (the wave front) there is a difference in path covered $u \sin \alpha$ between the two rays, and thus a phase difference $(2\pi u \sin \alpha)/\lambda$, when λ represents the wave length of the light used. If we represent the light vibration at A by $\sin\omega t$ (ω is the angular frequency of the light used), the light vibration at B varies according to $\sin(\omega t + 2\pi\alpha u/\lambda)$, where $\sin\alpha$ has been replaced by α since we have limited the case to very small angles of diffraction. When we pass on to the two-dimensional case, where the influence of the vertical component v and the vertical angle of diffraction β are taken into account, it is found that the ray α, β from the hole u, v with the dimensions du and dv may be represented in a corresponding way by

$$\sin\left(\omega t + \frac{2\pi\alpha}{\lambda}u + \frac{2\pi\beta}{\lambda}v\right) du\, dv \quad . \quad . \qquad (1)$$

When a positive lens is placed behind the film (*fig. 5*) every point in the focus corresponds to a direction α, β of diffraction. The coordinates of this focus are therefore equal to α and β, respectively,

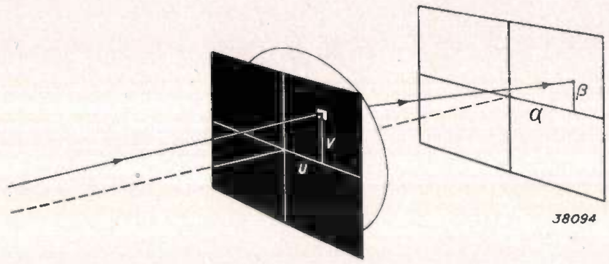


Fig. 5. Schematic representation of the arrangement for diffraction experiments. The strip of film is illuminated with a parallel beam of light. A lens is placed behind the film which concentrates all the rays diffracted by the same angles α, β in one point of the focal plane. There is assumed to be a rectangular hole in the film with the coordinates u, v.

except for a factor. According to (1) the focal plane is homogeneously illuminated, since the amplitude of the light oscillation is independent of α and β; the phase, however, contains α and β, and therefore differs from point to point.

Let us now imagine on the film an infinitely long horizontal row of holes at equal intervals l with the coordinates $u = x$, $x + l$, $x \pm 2l$, etc., see *fig. 6a*. On the focal plane the contribution of the m^{th} hole to the light oscillation is

$$\sin\left\{\omega t + \frac{2\pi\alpha}{\lambda}(x + ml) + \frac{2\pi\beta}{\lambda}v\right\} dx\, dv \quad . \quad . \qquad (2)$$

The total oscillation is then obtained by summation over all the holes. Physically this means an inter-

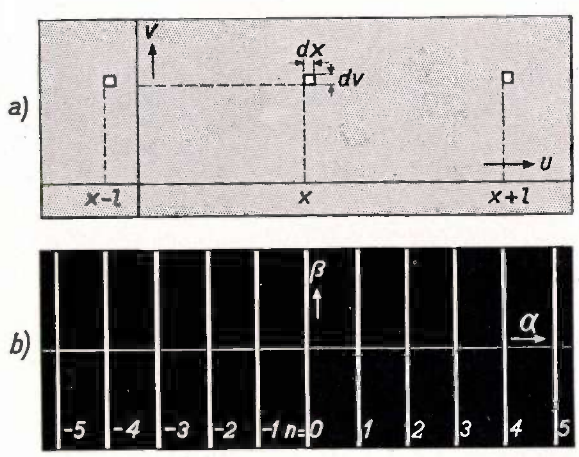


Fig. 6. A perforated grating (a), consisting of an infinite row of holes at height v and at distances l apart, causes a diffraction pattern which contains an infinite number of equidistant vertical lines (b) along which the light has a constant amplitude.

---

[6]) Mathematically speaking, the calculation of the non-linear distortion has of course nothing to do with the light diffraction. The method should only be considered as a welcome aid in bringing the very abstract manner of calculation (namely with the help of two-dimensional Fourier analysis) into expression in a graphic manner.

ference between rays having the same direction and coming from different holes. For certain directions, namely for values of a for which

$$a_n = \frac{n\lambda}{l}, \text{ mit } n = 0, \pm 1, \pm 2, \ldots, \quad (3)$$

the phase difference between each pair of rays will be an integral multiple of 2π, independent of m, whereby all the rays reinforce each other; for all intermediate directions the rays will cancel each other. The focal plane is therefore no longer homogeneously illuminated, the only light which remains is at a series of infinitesimally narrow vertical lines (fig. 6*b*).

For the amplitude of the light oscillation at these lines we would find infinity upon summation over the infinite number of holes. In the following, however, we shall consider the contribution per unit length of the film [7]). At the n^{th} line the contribution of each hole is equal to

$$\sin \left\{ \omega t + \frac{2\pi a_n}{\lambda} x + \frac{2\pi\beta}{\lambda} v \right\} \mathrm{d}x \, \mathrm{d}v,$$

per unit length of the film therefore

$$L_n = \frac{1}{l} \sin \left\{ \omega t + \frac{2\pi a_n}{\lambda} x + \frac{2\pi\beta}{\lambda} v \right\} \mathrm{d}x \, \mathrm{d}v \; . \quad (4a)$$

or, when (3) is substituted and when for the sake of simplicity we set $2\pi\beta/\lambda = b$,

$$L_n = \frac{1}{l} \sin \left\{ \omega t + \frac{2\pi n}{l} x + bv \right\} \mathrm{d}x \, \mathrm{d}v \; . \; . \quad (4b)$$

If, as will be the case in the following, we confine ourselves to the calculation of periodic gratings, we may always consider them to be made up of the sum of a large number of perforated gratings of the same period, shifted with respect to each other in horizontal and vertical direction. Each of these perforated gratings, independent of its position x, y, leads to the same set of equidistant lines (3), with, however, a difference in phase. The resulting periodic grating, no matter what complex form it may have in every period, will therefore only be able to exhibit intensities on the same vertical lines, and will only be distinguished by different intensity distributions in a vertical direction.

As a following step we shall now derive the diffraction pattern of the slit grating.

[7]) In practice one is always concerned with a finite length of film. This is manifested (loc. cit.[3,4]) solely in a slight broadening of all the lines in the diffraction pattern, but leaves the macro structure of the pattern unaltered.

The diffraction pattern of the slit grating

A grating consisting of equidistant vertical slits which extend from $v = 0$ to $v = y$ (*fig. 7a*) may be considered as the sum of perforated gratings placed

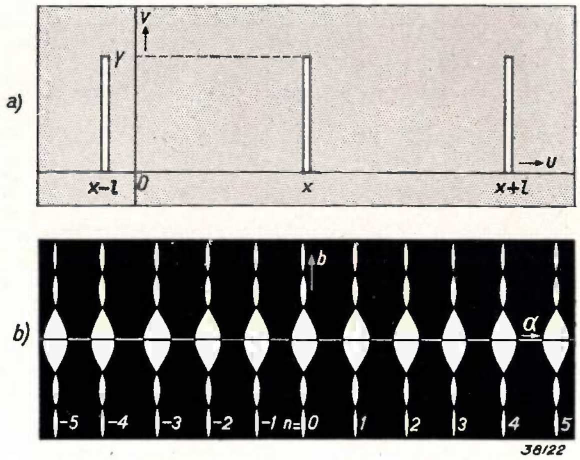


Fig. 7. A slit grating (*a*) with a height y and the period l causes a diffraction pattern (*b*) which consists of vertical lines in the same positions as in fig. 6*b*. Along each line, however, the amplitude of the light now varies periodically. This is indicated in the drawing by a variation of the thickness of the line.

vertically one above the other. The diffraction pattern obtained is the sum of the patterns of the individual perforated gratings. It is obtained by the integration of (4b) with respect to dv:

$$L_n = \frac{1}{l} \, \mathrm{d}x \int_0^y \sin\left(\omega t + \frac{2\pi n}{l} x + bv\right) \mathrm{d}v$$

$$= \frac{1}{bl} \left\{ \cos\left(\omega t + \frac{2\pi n}{l} x\right) - \cos\left(\omega t + \frac{2\pi n}{l} x + by\right) \right\} \mathrm{d}x \quad . \; (5a)$$

$$= \frac{2}{bl} \sin \frac{by}{2} \cdot \sin\left(\omega t + \frac{2\pi n}{l} x + \frac{by}{2}\right) \mathrm{d}x \; . \; . \; . \; . \; . \; . \; . \; (5b)$$

For each line n the amplitude of the light oscillation as a function of the vertical coordinate b is therefore given by the same function

$$\frac{2}{bl} \cdot \sin \frac{by}{2} \, \mathrm{d}x.$$

The amplitude [8]), which in the case of the perforated grating was constant over a whole line, has now become periodic along each line and the intensity of the maxima decreases proportionally to b (fig. 7*b*).

The diffraction pattern of a sinusoidal track. Unilateral modulation

If we now join slit gratings of different height y (but with the same period l) together in a horizon-

[8]) In visual or photographic observation of the spectrum one will always be concerned with the intensity of the light, which is proportional to the square of the amplitude of the light. For our considerations, however, it suffices if we deal with the amplitude in all cases.

tal direction, we obtain what is called in sound film technology a modulated sound track. For a pure tone

$$y = p + q \cos \frac{2\pi x}{l}, \quad \ldots \quad (6)$$

where p represents the width of the "zero track", and q the amplitude of the signal (*fig. 8*); 100 q/p is the percentage depth of modulation. In this way a so-called unilateral modulation has been obtained, which indicates that only one boundary of the zero track is varied.

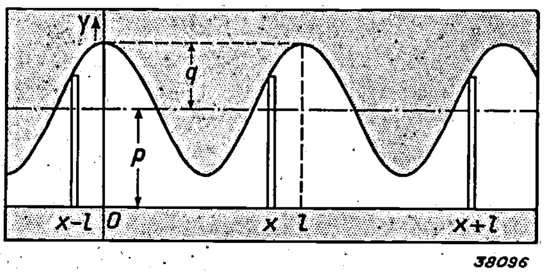


Fig. 8. A sound track containing a unilaterally modulated sine may be built up of a series of slit gratings, all of which have the same period l, but which are shifted with respect to each other in the horizontal direction. The zero track has the width p, the amplitude of the sine curve is q.

If we substitute condition (6) in (5a), and integrate with respect to dx, we obtain

$$L_n = \frac{1}{bl}\int_0^l \cos\left(\omega t + \frac{2\pi n}{l}x\right) \mathrm{d}x -$$

$$-\frac{1}{bl}\int_0^l \cos\left\{\omega t + \frac{2\pi n}{l}x + b\left(p + q\cos\frac{2\pi x}{l}\right)\right\} \mathrm{d}x.$$

The second integral upon working out contains terms of the form

$$\cos\,(bq \cos 2\pi x/l) \text{ and } \sin\,(bq \cos 2\pi x/l),$$

which cannot immediately be reduced to goniometric functions, but which may be expressed in a series with Bessel functions:

$$\cos\,(\xi \cos \Theta) = \sum_{k=-\infty}^{k=+\infty} (-1)^k \cos 2k\Theta \cdot J_{2k}\,(\xi),$$

$$\sin\,(\xi \cos \Theta) = \sum_{k=-\infty}^{k=+\infty} (-1)^k \cos\,(2k+1)\Theta \cdot J_{2k+1}\,(\xi).$$

By this means the expression for the light oscillation in the diffraction spectrum may be reduced to

$$L_n = -\frac{1}{b}\cos\left(\omega t + bp + \frac{n\pi}{2}\right) J_n\,(bq) \quad . \quad (7a)$$

for $n = \pm 1, \pm 2, \pm 3, \ldots$, while the zero order is given by

$$L_0 = \frac{1}{b}\cos \omega t - \frac{1}{b}\cos\,(\omega t + bp)\,J_0\,(bq) \quad . \quad . \quad (8a)$$

In these formulae J_n represents the Bessel function of the n^{th} order.

For a good understanding of the character of the diffraction spectra we shall for a moment consider the behaviour of these functions. They may be developed as follows in exponential series:

$$J_0(\xi) = 1 - \frac{\xi^2}{2^2} + \frac{\xi^4}{2^2{\cdot}4^2} - \frac{\xi^6}{2^2{\cdot}4^2{\cdot}6^2} + \ldots = +J_0\,(-\xi),$$

$$J_1(\xi) = \frac{\xi}{2} - \frac{\xi^3}{2^2{\cdot}4} + \frac{\xi^5}{2^2{\cdot}4^2{\cdot}6} - \ldots = -J_1\,(-\xi) = -J_{-1}\,(\xi),$$

$$J_2(\xi) = \frac{\xi^2}{2{\cdot}4} - \frac{\xi^4}{2^2{\cdot}4{\cdot}6} + \frac{\xi^6}{2^2{\cdot}4^2{\cdot}6{\cdot}8} - \ldots = +J_2\,(-\xi) = +J_{-2}\,(\xi),$$

$$J_3(\xi) = \frac{\xi^3}{2{\cdot}4{\cdot}6} - \frac{\xi^5}{2^2{\cdot}4{\cdot}6{\cdot}8} + \ldots = -J_3\,(-\xi) = -J_{-3}\,(\xi),$$

$$J_4(\xi) = \frac{\xi^4}{2{\cdot}4{\cdot}6{\cdot}8} - \frac{\xi^6}{2^2{\cdot}4{\cdot}6{\cdot}8{\cdot}10} + \ldots = +J_4\,(-\xi) = +J_{-4}\,(\xi),$$

usw.

As shown by the series and *fig. 9*, $J_0(\xi)$ shows great similarity to a cosine, although with gradually decreasing amplitude. $J_1(\xi)$ has the character of a sine, while the Bessel functions of higher orders are also oscillating functions but with progressing retardation as concerns their commencement. They reach their first maximum at a value of the argument which is greater than the order n by about one.

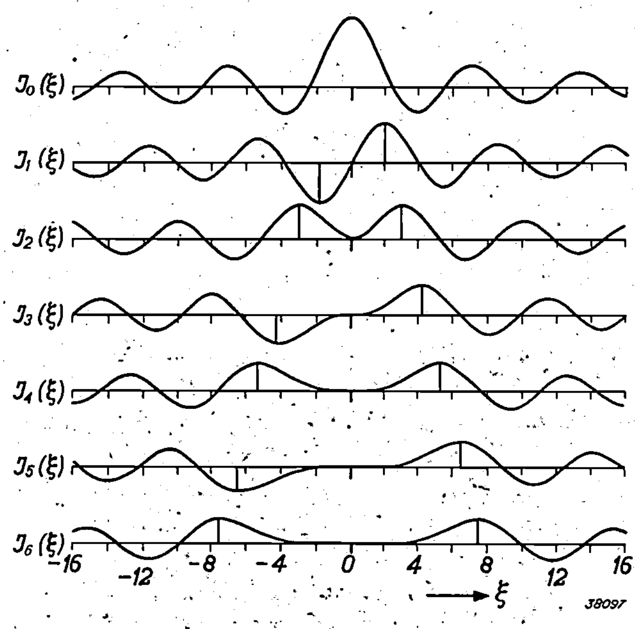


Fig. 9. Graphical representation of the Bessel functions ($J_n(\xi)$) for the orders $n = 0$ to $n = 6$.

According to (7a) and (8a) the intensity distribution along the line of n^{th} order in the diffraction pattern is determined by the Bessel function of that order. More exactly: the amplitude of the light is given by the quantity

$$\left|\frac{J_n\,(\xi)}{\xi}\right|$$

This quantity is represented graphically in the model of *fig. 10* by a number of positive and negative integral orders (except $n = 0$). The behaviour of the functions corresponds in character

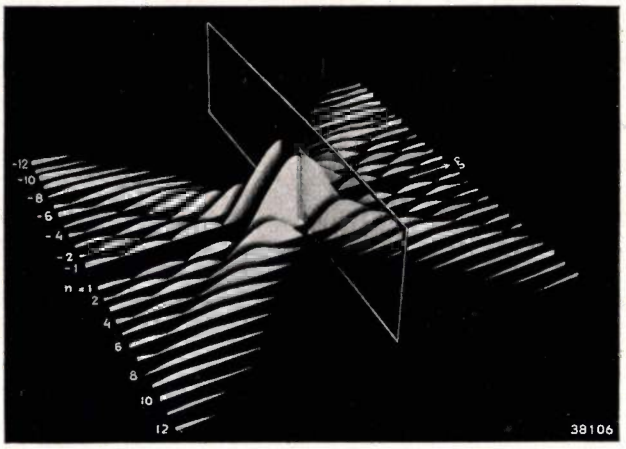


Fig. 10. Model of the function $J_n(\xi)$ for a number of whole number orders n, except zero. The vertical celluloid screen represents the plane through the axis $\xi = 0$. In this plane the function differs from zero only for $n = \pm 1$.

to that of the Bessel functions themselves (fig. 9), and this behaviour is found to be perfectly reflected in the spectrum recorded, *fig. 11*. It may be seen that the higher orders reach their first maximum at steadily increasing values of b, which gives the whole pattern the form of a St Andrew's cross.

For the amplitude on the horizontal axis ($b = 0$) it follows from (7a) and (8a) that:

$$A_0 = p$$
$$A_1 = q/2$$
$$A_2 = A_3 = A_4 = \ldots = 0.$$

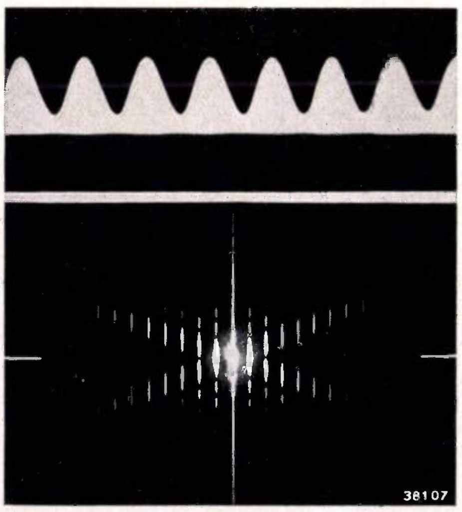


Fig. 11. Photograph of the diffraction pattern of the unilaterally modulated sine ($\nu = 400$ c/s). The light distribution in this pattern is given by the model shown in fig. 10. The horizontal axis $b = 0$ corresponds to the celluloid plane $\xi = 0$.

This confirms the proposition made at the beginning in the general considerations that the diffraction pattern on the horizontal axis provides the Fourier analysis of the sound: the zero order is determined exclusively by the width q of the zero track, the first order exclusively by the amplitude q of the signal, while all the higher orders have a zero intensity.

The diffraction pattern of a sinusoidal track. Bilateral modulation

At the present time it is almost universally customary in sound film technology to vary not one, but both boundaries of the zero track simultaneously. We shall derive the diffraction pattern so obtained from that of the unilaterally modulated track, and for this purpose we shall consider the bilaterally modulated track to consist of two parts: the part above and the part below the x-axis, *i.e.* two unilaterally modulated tracks. The upper part furnishes the pattern which we have already derived, the lower part is obtained by not integrating from 0 to y in the calculation of (5), but from $-y$ to 0. This comes down to substituting in formulae (7a) and (8b) $-p$ and $-q$ for p and q, and giving the whole a negative sign. This gives:

$$L_n' = +\frac{1}{b}\cos\left(\omega t - bp + \frac{n\pi}{2}\right) J_n(-bq), \quad (7b)$$

$$L_0' = -\frac{1}{b}\cos\omega t + \frac{1}{b}\cos(\omega t - bp)\, J_0(-bq). \quad (8b)$$

If the two patterns are now added together we obtain for all orders the light oscillation:

$$L_n = \frac{2}{b}\sin\left(bp + \frac{n\pi}{2}\right) J_n(bq)\sin\omega t \quad (9)$$

The distinction between the diffraction pattern now obtained and the original one consists in the fact that the width p of the zero track, which with unilateral modulation (see 7*a*) occurs only in the phase of the light oscillation (except in the zero order with which we are not further concerned), now enters the amplitude, and is therefore expressed in the intensity distribution. It causes thereby, as fig. 2 also shows, a series of equidistant zero points and maxima in each line, which alternate for neighbouring orders, since according to (9) they are determined for even orders by sin bp and for odd orders by cos bp.

We must therefore conceive these extra zero points and maxima as the result of the interference of the two diffraction spectra, which are projected by the upper and lower parts, respectively, of the track.

This phenomenon permits us to determine the quantities p and q separately from the spectrum, since they are present independent of each other. If the width of the zero track is changed with q remaining the same, the distance between the p zero points changes, while the St Andrew's cross retains its shape. If the depth of modulation is changed while the width of the zero track remains the same, the slope of the arms of the cross changes, while the p zero points retain the same separation.

On the horizontal axis, according to the above-mentioned proposition, we may again only expect the zero and first order. We find

$$A_0 = 2p$$
$$A_1 = q$$
$$A_2 = A_3 = A_4 = \dots = 0.$$

The diffraction pattern of obliquely recorded film

No matter how much the sound track may be deformed by oblique recording, it still remains periodical with the same period l. As we have explained in the discussion of the perforated grating, the diffraction pattern will therefore consist solely of vertical lines in their old positions α_n. No new lines occur, only the intensity distribution of each line in the vertical direction will be changed.

Oblique recording in our method of building up the sound track amounts to the fact that we must begin with a slit grating whose slits make a certain angle ψ with the vertical axis [9].

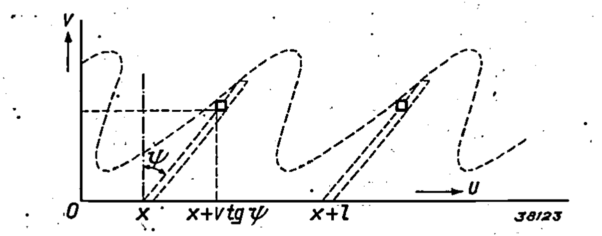


Fig. 12. An obliquely recorded sound track may be built up of gratings of oblique slits (angle ψ). The diffraction pattern of such a slit grating can be derived in the same way as above, by integrating over perforated gratings. The perforated gratings lying one above the other must then be shifted with respect to each other according to the condition $\mu = x + v \tan \psi$.

We now attempt to derive the diffraction pattern of this slit grating from that of the perforated grating. The perforated grating at the height v must now also be provided with another u coordinate, namely by the replacing of x by $x + v \tan \psi$ (see *fig. 12*). The light oscillation (4a), when we indicate the coordinates of the new diffraction spectrom by α' and β', thus becomes

$$\frac{1}{l} \sin \left\{ \omega t + \frac{2\pi\alpha'}{\lambda} x + \left(\frac{2\pi\alpha'}{\lambda} \operatorname{tg} \psi + \frac{2\pi\beta'}{\lambda} \right) v \right\} \mathrm{d}x \, \mathrm{d}v$$

or

$$\frac{1}{l} \sin \left\{ \omega t + \frac{2\pi\alpha}{\lambda} x + \frac{2\pi\beta}{\lambda} v \right\} \mathrm{d}x \, \mathrm{d}v, \quad \text{(4c)}$$

where

$$\left. \begin{aligned} \alpha &= \alpha', \\ \beta &= \beta' + \alpha' \operatorname{tg} \psi. \end{aligned} \right\} \quad (10)$$

With this the solution of our problem is already complete. Equation (4c) is absolutely identical with equation (4a), so that with it, upon passing on to the slit, the sine, etc., we derive exactly the

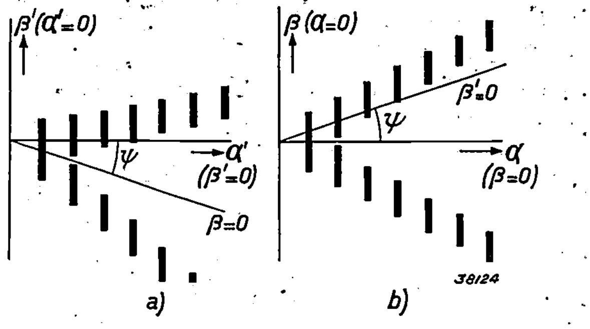


Fig. 13. Transformation experienced by the diffraction pattern when the sound track is recorded obliquely. All the vertical lines in the pattern remain vertical, all the horizontal lines are rotated through an angle $\psi(\alpha)$. In order to find the new acoustic spectrum, therefore, the old diffraction pattern (with vertical slit) may be used by rotating the "horizontal axis" through an angle ψ (b).

same diffraction patterns as for the vertically recorded films. The only difference is that the coordinates α and β must be replaced by α' and β' by the use of equation (10), which amounts to an oblique angular transformation of the normal diffraction pattern, in which the vertical lines remain vertical and the horizontal lines are rotated through an angle ψ (*fig. 13*). With this therefore we have deduced the behaviour of the diffraction pattern of obliquely recorded film already described in our general consideration.

This may be further illustrated by means of the model of the function $|J_n(\xi)/\xi|$, which was reproduced in fig. 10. The different functions which are cut out of paper are mounted in separate wooden strips. The spectrum of the obliquely recorded film is now obtained (*fig. 14*) by sliding the strips along each other. It is clear how the second and third orders now slide across the horizontal axis (the celluloid plane) and thus appear in the Fourier spectrum of the sound obtained upon reproduction.

The diffraction pattern of the normal true sine, which until now was of no acoustic interest outside the horizontal axis, now takes on full significance, since it is to be considered as a complete model of the Fourier spectra which will be obtained at any given angle of recording or reproduction. The new "horizontal axis" $\beta' = 0$ is found in the normal dif-

[9]) The obliquely recorded film may be considered to be formed from the vertically recorded film by an oblique angular transformation in which the horizontal lines have remained horizontal and the vertical lines have been rotated through an angle ψ.

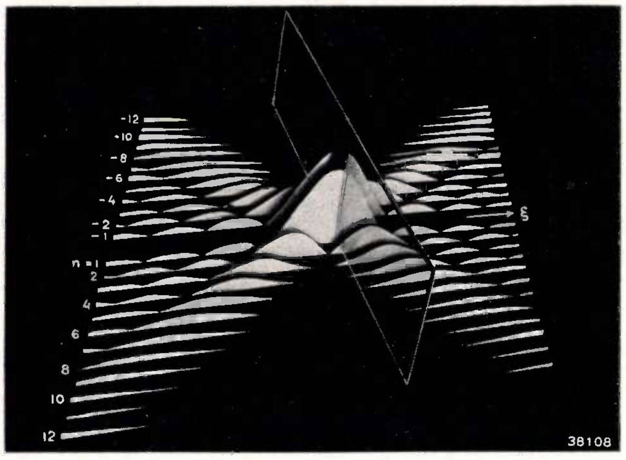


Fig. 14. The model shown in fig. 10 is so constructed that the strips of paper which represent the light distribution along the lines of the diffraction pattern can be shifted along each other. In this way the transformation shown in fig. 13 can be demonstrated.

fraction pattern according to (10) as a straight line through the origin

$$\beta = a \operatorname{tg} \psi \quad . \; . \; . \; . \; . \; . \quad (11)$$

The values of the amplitude of the light of the different orders on this line give us the coefficients of the different higher harmonics.

We shall now carry out this calculation numerically for several practically important cases.

The Fourier spectrum upon oblique recording or reproduction

Unilateral modulation

The amplitude of the light oscillation upon diffraction by a unilaterally modulated sine track is, according to (7a) proportional to

$$A_n = \frac{1}{b} J_n(bq) \quad . \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

If in (11) we again write $2\pi\beta/\lambda = b$ and $a = n\lambda/l$ and express the period l in the frequency ν and the film velocity c ($c = \nu l$), (11) becomes

$$b = \frac{2\pi n\nu}{c} \operatorname{tg} \psi \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (13)$$

For small angles ψ, thus for small values of bq, in the series of the Bessel functions the first term is sufficient, and tan ψ may be replaced by ψ. We then obtain

$$A_n = \frac{b^{n-1}\, q^n}{2^n \cdot n!} = \frac{q}{2 \cdot n!} \left(\frac{\pi n q \nu \psi}{c}\right)^{n-1},$$

thus for the separate amplitudes A_n and for the relative amplitudes B_n, referred to the fundamental frequency:

$$A_1 = \frac{q}{2}, \qquad B_1 = \frac{A_1}{A_1} = 1,$$

$$A_2 = \frac{\pi q^2}{2c} \nu\psi, \qquad B_2 = \frac{A_2}{A_1} = \frac{\pi q}{c} \nu\psi,$$

$$A_3 = \frac{3}{4} \frac{\pi^2 q^3}{c^2} \nu^2\psi^2, \qquad B_3 = \frac{A_3}{A_1} = \frac{3}{2} \frac{\pi^2 q^2}{c^2} \nu^2\psi^2 = 1{,}5\; B_2^{\,2},$$

$$A_4 = \frac{4}{3} \frac{\pi^3 q^4}{c^3} \nu^3\psi^3, \qquad B_4 = \frac{A_4}{A_1} = \frac{8}{3} \frac{\pi^3 q^3}{c^3} \nu^3\psi^3 = 2{,}33\, B_2^{\,3}.$$

It may therefore be seen that the amplitude A_1 of the fundamental frequency is independent in the first approximation of the frequency ν and the angle ψ. The relative intensity B_2 of the second harmonic is directly proportional to the sound amplitude q and to ν and ψ, the third harmonic to the square of these quantities, etc. We are therefore concerned with a nonlinear distortion which increases rapidly with the frequency. This distortion is independent of the width p of the zero track. For small angles the higher harmonics may be neglected compared with the second harmonics, so that we need only deal with the latter.

If we assume a film velocity of 320 mm/sec and an amplitude q of 1.0 mm, and if we express the frequency ν in kc/s and the angle ψ in minutes,

$$B_2 = 0{,}29\, \nu\psi\; \% \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (14)$$

The fundamental frequency is indeed independent of ψ in the first approximation used until now, but its amplitude will decrease in the second approximation as the model (fig. 14) also shows. Here also the quantity $q\nu\psi$ is the determining factor. In sound film practice therefore the correct position of the light slit is checked by recording the frequency 8 000 c/s and by adjusting the slit for reproduction so that a maximum sound intensity is obtained. The accuracy of this adjustment, particularly in unilateral modulation, may be seen from the following example. It is for example required that the second harmonic should not be stronger than 3 per cent at 4 000 c/s. This means according to (14) an angle ψ of $2^1/_2$ minutes. From *fig. 15*, curve *a*, it then follows that upon adjustment with the frequency 8 000 c/s the amplitude of the fundamental frequency may not deviate more than 0.15 per cent from the maximum value.

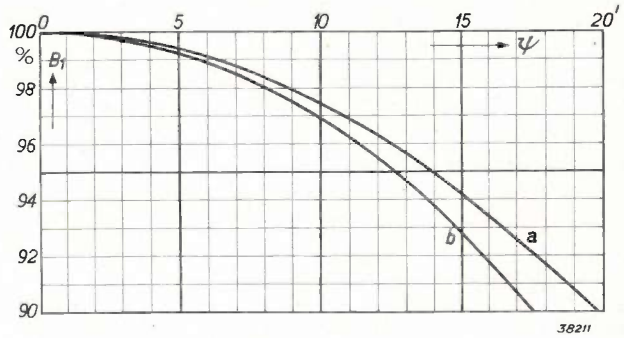


Fig. 15. Amplitude of the fundamental tone of the sound which is obtained upon oblique scanning of a sinusoidal sound track, as a function of the angle ψ of the slit. The curve *a* is for unilateral, *b* for bilateral modulation (frequency ν = 8 000 c/s, depth of modulation 100%, zero track width 1.0 mm, film velocity 320 mm/sec).

Bilateral modulation

The amplitude of the light oscillation upon diffraction by a bilaterally modulated sine track is, according to (9),

$$A_n = \frac{2}{b} \sin\left(bp + \frac{n\pi}{2}\right) J_n(bq).$$

For odd orders this amplitude is the same as the one just dealt with, because for small angles cos bp may be considered equal to unity. For even orders, however, a multiplication factor sin bp occurs, so that the zero track width p now begins to play a part. For small angles ψ we find

$$A_1 = q, \qquad B_1 = \frac{A_1}{A_1} = 1,$$
$$A_2 = 4\,\frac{\pi^2 pq^2}{c^2}\,\nu^2\psi^2, \qquad B_2 = \frac{A_2}{A_1} = 4\,\frac{\pi^2 pq}{c^2}\,\nu^2\psi^2,$$
$$A_3 = \frac{3}{2}\,\frac{\pi^2 q^3}{c^2}\,\nu^2\psi^2, \qquad B_3 = \frac{A_3}{A_1} = \frac{3}{2}\,\frac{\pi^2 q^2}{c^2}\,\nu^2\psi^2 = 0{,}375\,\frac{q}{p}\,B_2,$$
$$A_4 = \frac{64}{3}\,\frac{\pi^4 pq^4}{c^4}\,\nu^4\psi^4, \qquad B_4 = \frac{A_4}{A_1} = \frac{64}{3}\,\frac{\pi^4 pq^3}{c^4}\,\nu^4\psi^4 = 1{,}33\,\frac{q}{p}\,B_2^2.$$

The second harmonic has now become proportional to the square of the frequency ν and the angle ψ, and has therefore fallen to about the same level as the third harmonic. Bilateral modulation is therefore considerably more satisfactory than unilateral modulation [10]) as far as this distortion is concerned. The case may be so conceived, that the upper and lower part of the track are, as it were, in push-pull connection, so that the even harmonics are suppressed in the first approximation.

If we assume a total zero track width $2p = 1.0$ mm and a depth of modulation of 100 per cent

$$B_2 = 0{,}0009\,\nu^2\psi^2\,\% \quad \ldots \ldots \ldots \quad (15)$$

If we now again require a maximum distortion at 4 000 c/s of 3 per cent, the maximum permissible angle $\psi = 15$ minutes, which condition, according to fig. 15, curve b, can be satisfied by adjusting the amplitude of the fully modulated tone 8 000 within 7 per cent of the maximum.

In *fig. 16* the percentage of second harmonic is given as a function of the angle ψ and the frequency ν for unilateral and bilateral modulation.

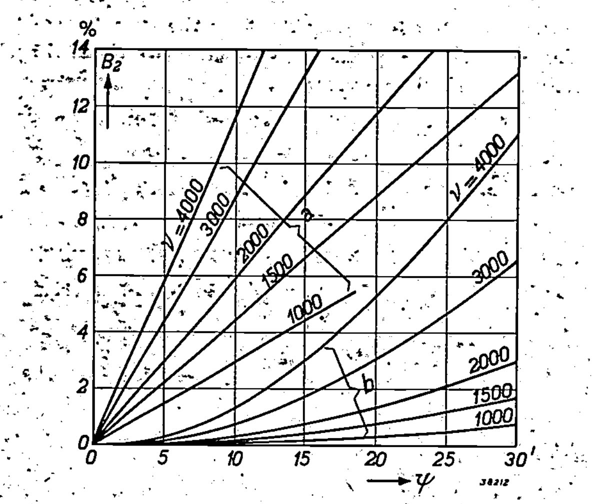


Fig. 16. Percentage B_2 of second harmonic as a function of the angle ψ of the slit for different frequencies ν. The group of curves a holds for unilateral, the group b for bilateral modulation (depth of modulation 100%, zero track width 1.0 mm, film velocity 320 mm/sec). It may be seen that in the latter case the reproduction is much less sensitive to slight deviations from the vertical position ($\psi = 0$) of the slit.

The diffraction pattern of two frequencies

When two frequencies ν and μ with amplitudes q and r, respectively, are recorded on the film, then the following is valid for y:

$$y = p + q\cos\frac{2\pi\nu}{c}\,x + r\cos\frac{2\pi\mu}{c}\,x,$$

where c is the film velocity in recording and reproduction. This also may always be interpreted as a periodic phenomenon, with a period which is equal to the smallest common multiple of the periods c/ν and c/μ. The calculations, which we shall not go into here [4]), may then take place quite analogously to the case of a single frequency. It is found that for angles α only lines can occur which satisfy the following:

$$\frac{c\,\alpha_{n,m}}{\lambda} = n\nu + m\mu, \quad \ldots \quad (16)$$

where n and m are positive or negative whole numbers. Besides the higher orders $n\nu$ and $m\mu$, therefore, all the intercombination lines occur.

With bilateral modulation the light oscillation is

$$L_{nm} = \frac{2}{b}\sin\left(bp + \frac{n+m}{2}\pi\right) J_n(bq)\,J_m(br)\,\sin\omega t \quad \ldots \quad (17)$$

This is therefore determined by products of Bessel functions.

In *figs. 17a* and *18a* two typical examples are reproduced. Fig. 17a shows the diffraction pattern of a sound track with two only slightly differing frequencies ($\nu = 1\,600$, $\mu = 1\,880$ c/s), fig. 18a the diffraction pattern for two frequencies far apart ($\nu = 1\,600$, $\mu = 210$ c/s). Figs. 17b and 18b show the diffraction patterns for the same combinations of frequencies, but in this case for obliquely recorded film. As was to be expected, the patterns have remained the same except for the oblique angular transformation. Since the normal pattern also contains all the intercombination lines outside the horizontal axis, they will occur in the sound upon reproduction with an oblique slit as combination tones, and to a greater degree the higher the order of the lines in question (the higher the frequency) and the closer to the horizontal axis they reach their first maximum.

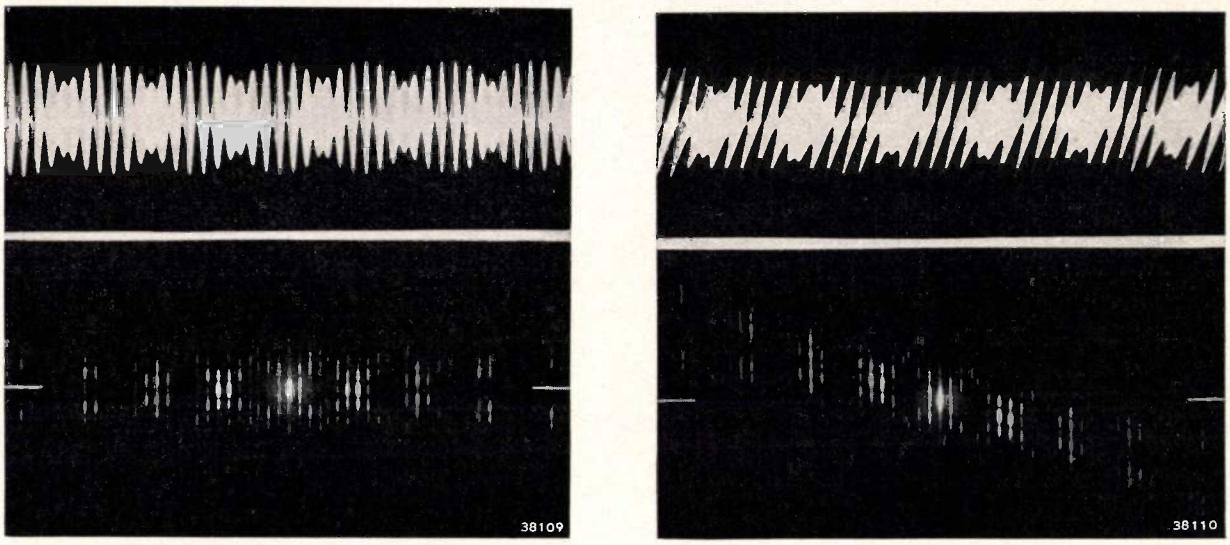
The calculation of this non-linear distortion is quite analogous to that with one frequency. We shall deal with one case as an example, namely the intensity of a low difference frequency of two high parent frequencies. The two high parent frequencies will be much distorted. What is now the magnitude of the difference frequency? Is it of the same order as the two second harmonics, or much smaller? The amplitude of the light oscillation in the diffraction spectrum, according to (17) is

$$A_{nm} = \frac{2}{b}\sin\left(bp + \frac{n+m}{2}\,\pi\right) J_n(bq)\,J_m(br).$$

The difference tone is obtained by setting $n = -1$

[10]) M. C. Batsel and E. W. Kellogg, J. Soc. Mot. Pict. Eng. 28, 516, 1937.

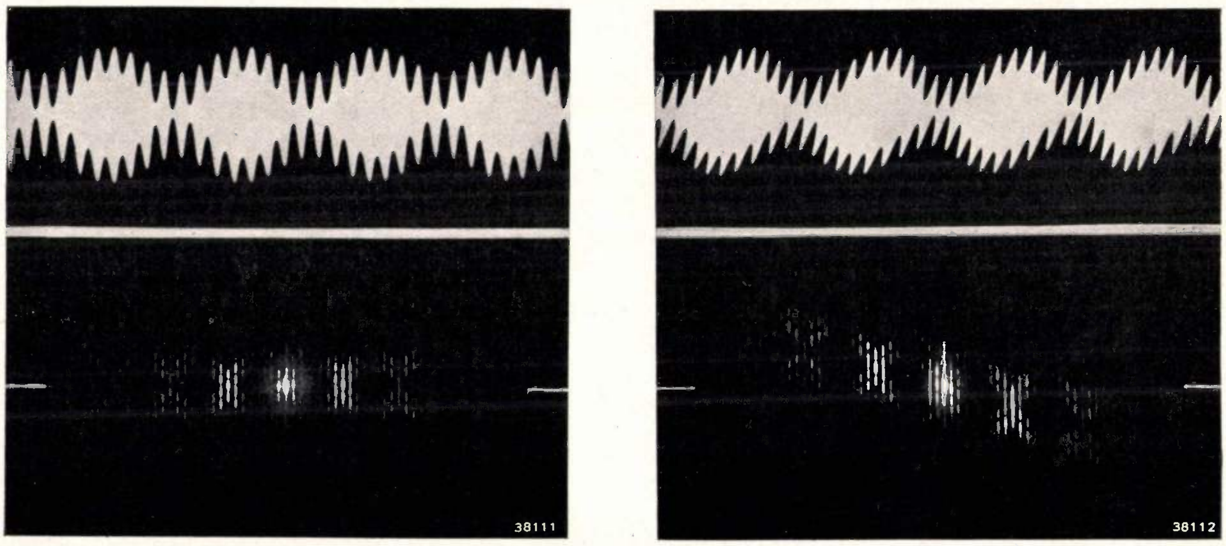
a



b

Fig. 17. Sound track with two different frequencies (ν = 1 600, μ = 1 880 c/s) and corresponding diffraction pattern, *a*) for recording with vertical slit, *b*) for recording with oblique slit (ψ = 15°). In (*a*) only the frequencies 0, 1 600 and 1 880 should be visible on the horizontal axis. The difference tone 280 is, however, also clearly visible, which indicates that a non-linear distortion was already present in the sound recorded.

a



b

Fig. 18. The same as in fig. 17 but for two frequencies lying far apart (ν = 1 600, μ = 210 c/s). Again *a*) was recorded with vertical slit and *b*) with oblique slit. Here also the recorded sound was already distorted, as may be seen from the fact that in (*a*) on the horizontal axis the sum and difference frequencies 1 810 and 1 390 and the second harmonic 3 200 are visible.

and $m = +1$, which gives as a first approximation:

$$A_{-1,+1} = -\frac{2}{b}\, bp\, \frac{bq}{2} \cdot \frac{br}{2} = -\frac{b^2 pqr}{2}\,.$$

Now according to equation (11) the value of b which corresponds to the intersection with the "horizontal axis" is always proportional to a, thus, according to (16) proportional to the difference frequency $\mu - \nu$ formed. The amplitude of the difference tone is therefore proportional to $(\mu - \nu)^2$.

We thus encounter a very remarkable kind of frequency-dependent non-linear distortion, in which the intensity of the combination frequencies formed is independent of the height of the parent frequencies and determined only by the height of the frequencies formed. If the difference frequency is low, it occurs only weakly, even though the two parent frequencies are both very much distorted.

A TELEPHONE INSTALLATION ON ULTRA SHORT WAVES FOR THE TROPICS

by C. G. A. von LINDERN.

621.396.5.029.6

For telephone communication in tropical and subtropical regions Philips have developed a transmitting and receiving installation which works on wavelengths of about 4 m. By the use of directional aerials a distance of from 50 to 100 km can be bridged with a transmitting power of 40 W if the transmitting and receiving installations are situated at sufficiently great altitudes, as on mountain tops. With this latter condition in view the installation is arranged for operation from a distance, while at the same time all precautions have been taken to protect the transmitter against the dangers involved in being set up in the open in a tropical climate. After a short description of the construction of the transmitter, the precautions mentioned are dealt with in detail. These are embodied in all kinds of structural particulars of the transmitter. In conclusion the operation from a distance of the installation is discussed.

For the telephonic bridging of oceans or extensive regions which are thinly populated it is usually simpler to achieve a radio connection than a cable connection; with very great distances the radio connection is indeed the only possibility. But even when it is a question of short distances, of only 50 or 100 km, for instance, under certain circumstances a radio connection may be preferable to a cable connection. Such conditions occur especially in tropical and subtropical regions: jungle, swamps and deserts in this case seriously hamper the installation of overhead wiring or the laying of cables, while, moreover, the cable connection always remains very vulnerable due to climatological conditions, and is accessible only with difficulty for any necessary repairs.

For such cases a radio connection on ultra short waves below 5 m is indicated. We shall describe an installation designed for this purpose, in which particular attention has been paid to the requirements made of the construction in a tropical climate.

The general design

The ultra short waves below 5 m have only a limited sphere of action with the output which may be used in practice: roughly speaking these waves are not propagated much farther than the distance of optical visibility. For the cases here under consideration, however, this is no objection, since it is only a question of relatively short distances which, when the transmitting and receiving aerials are mounted at great enough altitudes, make the two stations visible to each other. On the other hand the use of waves shorter than 5 m offers special advantages. In the first place in this wave-length region numerous transmitters may be used without their frequency ranges interfering with each other, an advantage which is further increased by the above-mentioned limitation of the sphere of action, since this makes it possible to use a given wave length several times in the same district. The great freedom in the frequency range can also be used to give the transmitter and receiver a band width such that a number of conversations can be transmitted simultaneously in different channels of the side bands (carrier-wave telephony). In the second place, with respect to transmitting energy also, short waves have advantages over longer waves. While it is indeed more difficult to generate large transmission outputs on short waves, there is the possibility of concentrating the energy emitted in a beam. At a wave length of 4 m with a directional aerial of reasonable dimensions a tenfold amplification of the energy radiated in a given direction can be obtained (of course at the expense of the energy radiated in other directions). For the case in question where the radio connections form part of a telephone network, and where, therefore, several of such radio connections must be connected in series, a signal intensity at the receiver must be required which is at least 50 dB above the normal interference level (noise). By the use of directional aerials (at the receiving end also) this signal to interference ratio, at a distance of 50 km, for example, and with the use of the best receiving valves available at the present time, can already be obtained with a transmission output of 30 to 40 W.

We have already mentioned that it is necessary to mount the aerials of transmitter and receiver at a great altitude, a mountain top for instance, in order to obtain the direct "line of vision". Since the supply line for the aerial must be short, the transmitter and receiver must be in the immediate neighbourhood of the aerials, while the telephone exchange where the radio connections terminate may in general be at some distance, in the valley for instance. This has two consequences: the transmitter and receiver must in general be set up in the open and work without attention, and the switching on of the apparatus must be able to take place from a distance. In the following there will be opportunity of discussing the consequences which these conditions lead to in the construction of the system.

In *fig. 1* the situation described is sketched simply. For transmitter and receiver directional aerials may be used which consist of a number of parallel rods mounted one behind the other. For the conversations in the two directions a slightly different wave length is used in order that the receiver of each station will not be affected by the transmitter of the same station.

achieved *via* the above-mentioned alternating magnetic field, see *fig. 2*. Only those oscillations with the wave length λ are then built up and maintained.

The resonance curve of the long line is sharper, *i.e.* the frequency of the transmitter is more constant, the fewer the dielectric and other losses occurring in the long line. The dielectric losses in our case are practically zero, since the concentric "cable" is built up entirely without solid insulating materials: the outer conductor is a strong copper cylinder 25 cm in diameter, the inner conductor a rod of 5 cm diameter and 1 m length inside the former, and entirely free of it. The losses by radiation

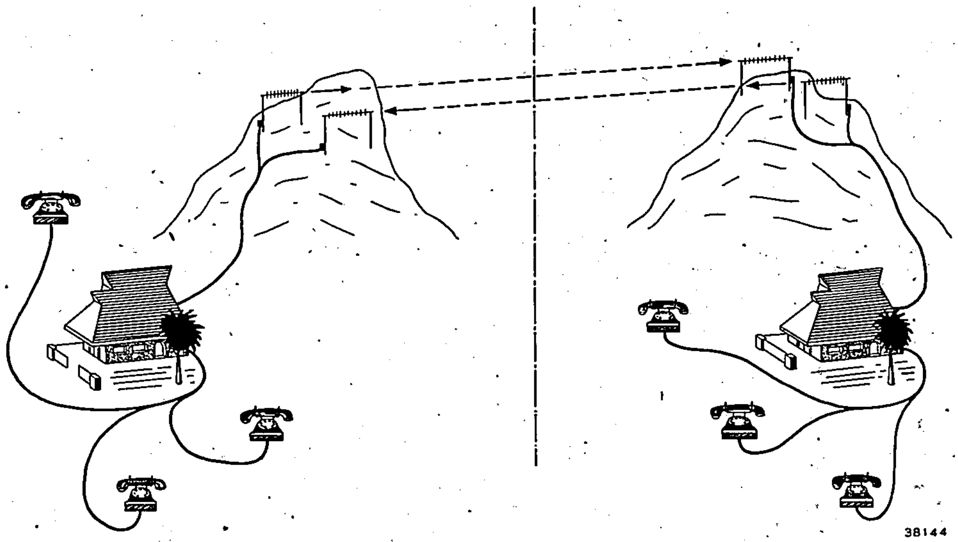


Fig. 1. Sketch of the arrangement for the radiotelephonic connection. In the neighbourhood of each of the two telephone exchanges to be connected the transmitter and the receiver with their directional aerials are set up at a high altitude. Each installation is connected to the corresponding telephone exchange by means of a telephone line and a power cable for the supply.

Connections of the transmitter

The transmitter proper consists of three stages: an oscillator stage which furnishes a high-frequency oscillation of small amplitude but with a constant frequency in the neighbourhood of 75 Mc/s (4 m wave length), an intermediate stage and an output stage in which the oscillation is amplified to an output of 40 watts. The frequency of the oscillator stage, which contains two pentodes PE 06/40 in push-pull connection, is determined by a so-called long line. This is a concentric cable short circuited at one end and open at the other. If the length of the inner conductor is l, with oscillations of the wave length $4l$, or $4l/3$ or $4l/5$ etc. resonance occurs, *i.e.* only for these wave lengths does an appreciable alternating magnetic field between the two conductors occur. The back-coupling of the anode circuit, roughly tuned to $\lambda = 4l$ by coils and condensers, on the grid circuit of the oscillator is now

and induction in the surroundings are rendered very small by making the outer conductor somewhat longer than the inner conductor and shutting it off completely (*fig. 3*).

Due to the coupling with anode and grid circuit of the oscillator valves, the characteristic frequency of the long line is to some degree affected by the rest of the connections. This undesired influence is kept small by not making the coupling too firm, and by including only a small anode impedance in the oscillator. This all means that only a small high-frequency output can be obtained. The amplification of the high-frequency oscillation in intermediate and output stage takes place with triodes TE 05/10 and TB 1/60, respectively, whereby the undesired reaction of the anode on the grid circuit caused by the anode-grid capacity is eliminated with neutrodyne condensers.

Modulation is accomplished by changing the

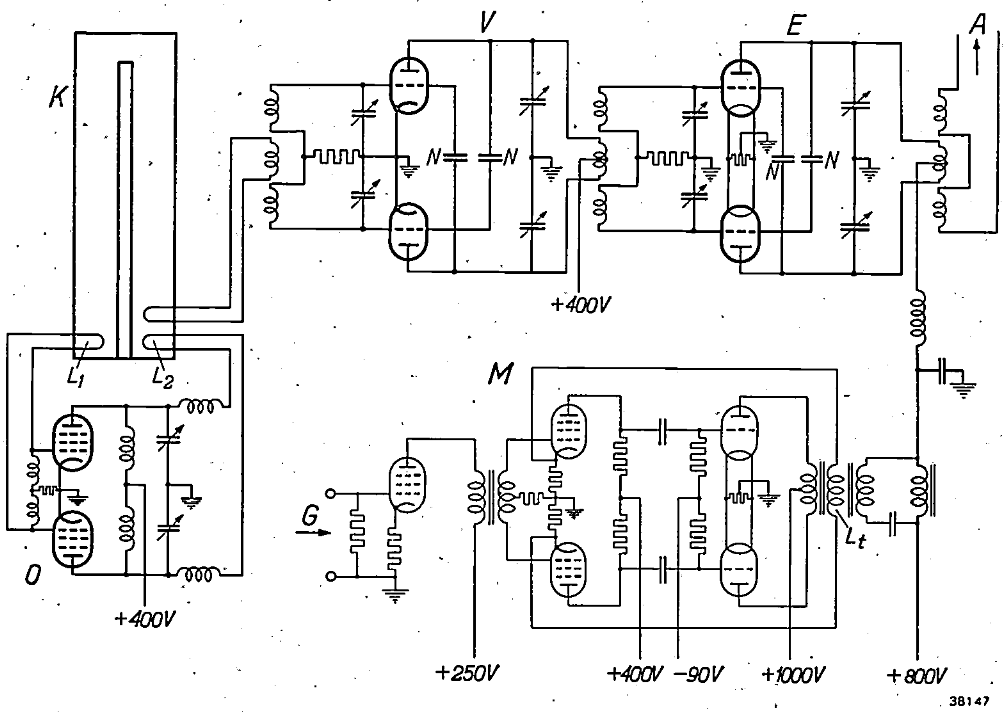


Fig. 2. Connections of the transmitter (simplified). The backcoupling of anode circuit on grid circuit of the oscillator O takes place through coils L_1, L_2, which are within the alternating magnetic field of the "long line" K. V intermediate amplifier stage, E output stage, N neutrodyne condensers, A aerial, M modulator. Modulation is accomplished by changing the anode voltage of the output stage. The modulating voltage entering at G is amplified in three stages; the coil L_t provides a tenfold inverse feed-back.

anode voltage of the final stage (fig. 2). The microphone voltages coming in over the line from the telephone exchange are amplified in three stages, so that it is possible to modulate the transmitter fully with a low-frequency input power of only $^1/_4$ milliwatt (an ordinary telephone apparatus provides 1 mW, so that the telephone line to the transmitter may cause an additional attenuation of 6 dB,

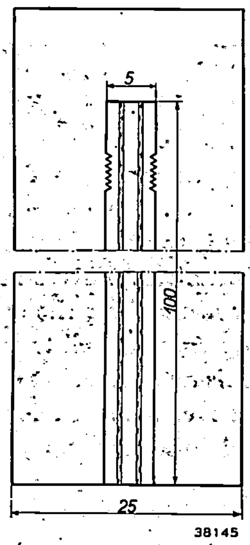


Fig. 3. Construction of the long line. It consists of a copper cylinder 25 cm in diameter in which a copper conductor 5 cm in diameter is fastened concentrically. The length of the inner conductor determines the wave length of the characteristic oscillation of the whole.

i.e. by a factor of 4). The frequency region of the modulator lies between 300 and 50 000 c/s for the first installation constructed, *i.e.* in this region the amplification factor remains constant within 2 dB, while the non-linear distortion with full load amounts to less than 1 per cent. This latter is obtained by a tenfold inverse feed-back. The broad frequency band, as was already mentioned at the beginning, permits telephony on a number of channels at the same time. If desired the modulator can be constructed for a still much wider frequency region having a width of several 100 kc/s without great difficulty.

As to the supply of the transmitter, it is assumed that there is a 220 volt A.C. main at the telephone exchange, to which the transmitter can be connected. The necessary rectifiers are therefore introduced into the transmitter in order to obtain the anode voltages, heating currents, etc. from the main. If no mains connection is available, as may sometimes be the case in places far from civilization, a petrol aggregate must be used for generating the current.

Construction of the transmitter

A transmitter which must function in a tropical climate and, moreover, in the open and without supervision, is exposed to many kinds of dangers.

It is protected against rain and sand by being enclosed in a strong tight case which is hung several metres above the ground on an aerial mast to remove

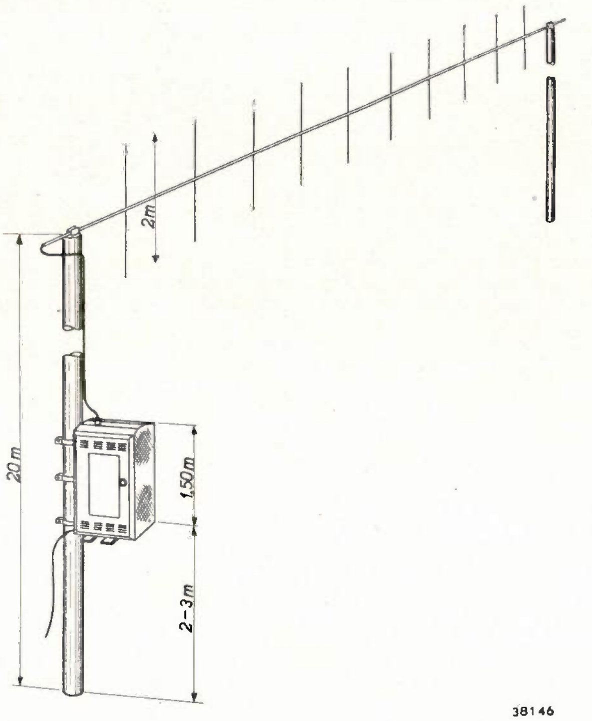
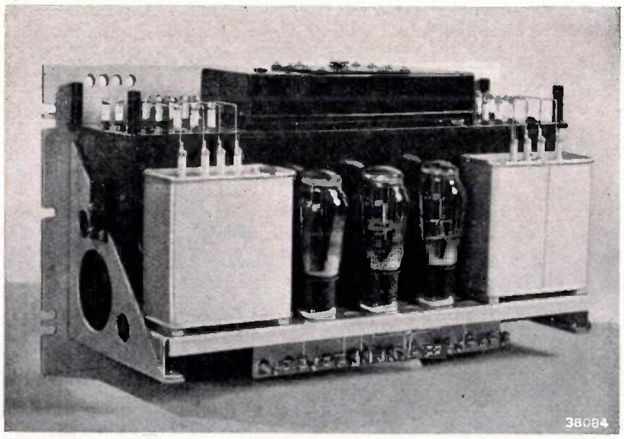
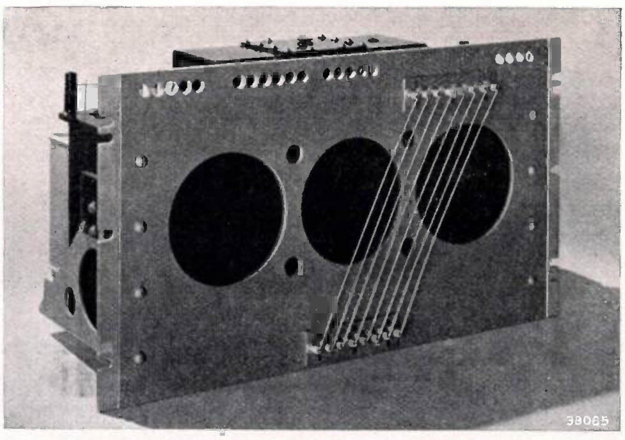


Fig. 4. The transmitter is enclosed in a strong closed case which is hung several metres above the ground on one of the 20 m high aerial poles.

it from the undesired attentions of the uninitiated and of all kinds of animals (*fig. 4*). A second danger ready to attack the transmitter are various kinds of fungi and the white ants so feared in the tropics. If it is not possible to seal the case by soldering it —which is of course the case here, since it must be opened once in a while for inspection and repair — there is only one sure method of preventing damage by these insects: in constructing the transmitter all materials must be avoided which could possibly stimulate the appetites of white ants and the like. This means in the first place that no textile or similar material may occur as insulation for the wiring; the wiring is thus bare and is supported by small insulators of high-frequency porcelain. Where "edible" insulation materials were unavoidable, as in the windings of the transformers, and chokes of the supply apparatus, the parts are housed in sealed oil containers with porcelain leads soldered in, see *fig. 5*. At the same time the temperature increase of the windings is hereby decreased which is doubly welcome in connection with the third great difficulty encountered in the tropics: the high temperature of the surroundings. This latter is particularly unpleasant, since the cooling of the transmittter is made difficult by the closed case which may not even contain any ventilation openings. In order to promote the dissipation of the heat developed in the case — the transmitter consumes 660 W — the walls and doors are made of double plates between which the outside air can circulate. In this way a larger surface takes part in the transmission of heat to the outside. In the case itself local overheating is avoided by a forced air circulation by means of fans. This necessitated a favourable arrangement of the different parts of the transmitter. In *fig. 6* the arrangement may be seen. The long line is placed vertically along one side, and above it the oscillator stage. In the middle of the case is a frame, to the front and back



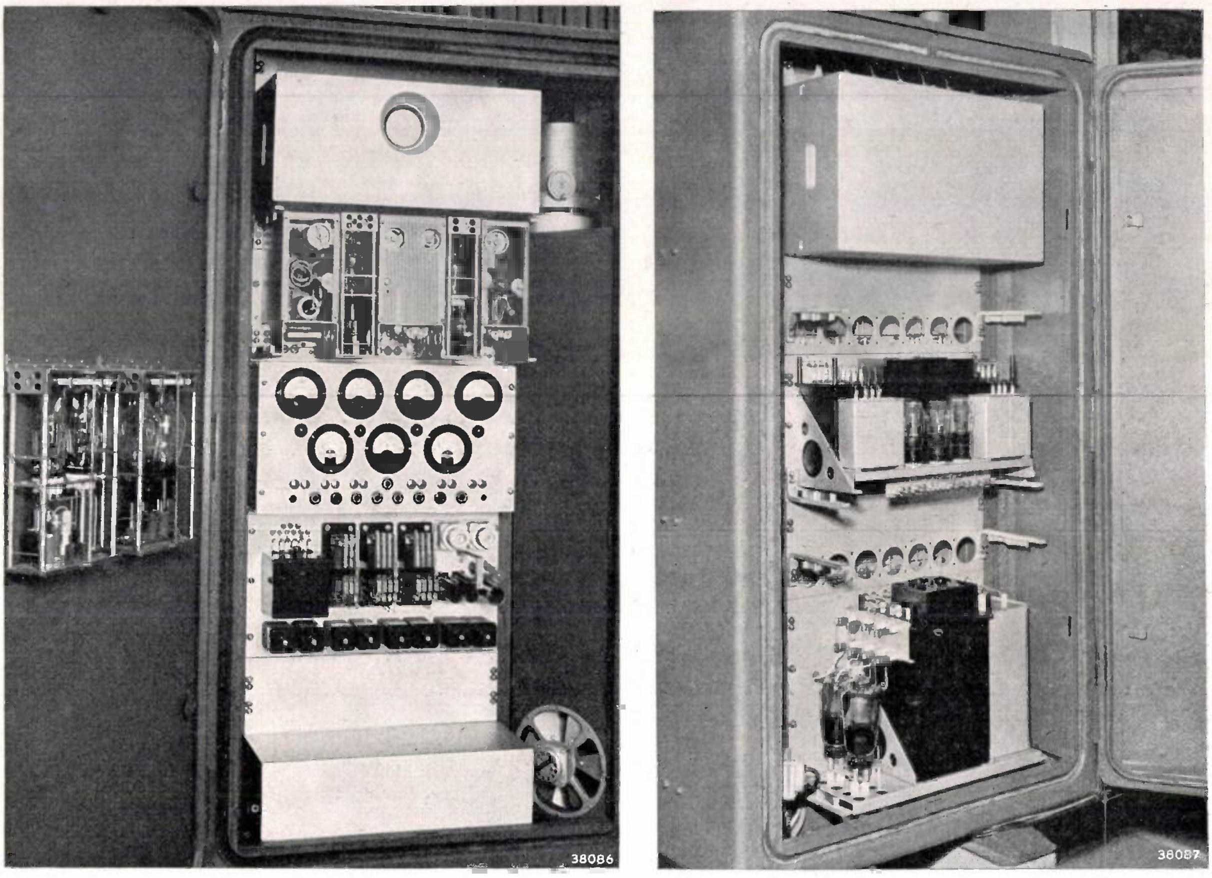
a



b

Fig. 5. Panel with supply apparatus *a*) front, *b*) back. The large container filled with oil contains the transformers and choking coils. On top of the container is an expansion chamber. Smoothing condensers are mounted in the boxes on either side of the rectifier valves. The wiring is bare, all leads through are made of high-frequency porcelain. The large holes in the rear wall of the panel promote the air cooling.

of which, both accessible by means of doors, a number of panels are fastened which contain the other stages of the transmitter, the supply apparatus, measuring instruments, switching arrangements, etc. The frame with its panels is separated from the long line by a partition; the fans which cause the air to flow upward along the panels and down along the long line are set into this partition. increase in temperature. If simply a copper rod with a coefficient of expansion of 162×10^{-7} were taken as inner conductor, upon a temperature change of 20° the length would vary by 0.032 per cent, which, with a wave length of 4 m would mean a shift of the carrier-wave frequency by about 24 000 c/s. In order to combat this very undesired effect the inner conductor is made of a tube of



a *b*

Fig. 6. Transmitter case with opened doors. At the front (*a*) may be seen to the right the partition which shuts off the long line and above the latter the oscillator stage. In addition a number of panels which contain (from top to bottom): a cathode ray tube, intermediate and output stage, measuring instruments, relays and switches. On the door are two reserve units. At the back (*b*) may be seen three panels (from top to bottom): the modulator, the supply apparatus for the heating voltages and the anode voltages of the first stages of transmitter and modulator, and the large supply apparatus for the anode voltage (1 000 volts) of the output stages of transmitter and modulator. Through the ventilation holes the wiring lying in the frame may be seen.

Thanks to these measures the temperature of all the parts is limited to a maximum of 70 °C at a temperature of the surroundings of 45 °C.

We shall now discuss several details of the construction.

The construction of the long line has already been discussed to some extent in the foregoing, and it was pointed out that the frequency is determined by the length of the inner conductor. Now this length changes due to the expansion of the material upon quartz over which a thin covering of copper is slid. The covering consists of two sections which are fastened to the ends of the quartz tube and connected with each other by an intermediate section which is extensible like an accordion, see fig. 3. The result is a copper rod whose length is as it were determined by the quartz tube. The coefficient of expansion of quartz is only 6×10^{-7}, so that the same calculation as above now gives a variation of only 900 c/s.

Since the grid-anode capacity of the triodes which are used in the intermediate and output stages of the transmitter may be somewhat different for different valves, upon changing defective valves the neutrodyne condensers must be adjusted anew. In order that this may also be done by unskilled personnel, two complete reserve units are hung on the inside of the door of the transmitter (see fig. 6*a*), which units are neutrodyned in advance, and can simply be slid into the place of the old units in the corresponding openings of the panel involved. The necessary connections are automatically made by a set of contacts. In *fig.* 7

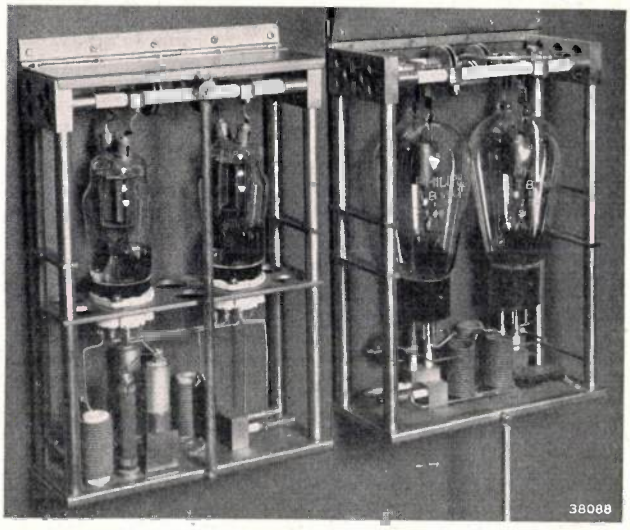


Fig. 7. Complete reserve units neutrodyned in advance. To the left for the intermediate stage, to the right for the output stage. These units can be slid into appropriate openings in the transmitter panel (fig. 6*a* second from the top).

the reserve units are shown separately. It may be seen that the connection between the valves and the other parts of the circuit have been kept very short, which is desirable in connection with the high transmission frequency. The neutrodyne condensers, which are clearly visible in the right-hand unit in fig. 7, consist of two plane parallel metal discs which are held apart by porcelain rods. The distances between the condenser plates, as well as between other points between which high-frequency voltages act [1]), are so chosen that even in mountainous districts, where due to the decreased atmospheric pressure and other causes the breakdown voltage may be considerably lower than at sea level, no breakdowns can occur.

The oil containers in which the transformers, etc. are housed have an expansion chamber in which the excess oil is driven when the liquid expands due to the heat (at 50° temperature difference the change in volume amounts to about 5 per cent). In order to prevent the oxidation of the oil as far as possible the expansion chamber is so constructed that the surface of contact between air and oil is small. Upon cooling, the oil runs back, and in doing so draws in a quantity of fresh air. Since in tropical climates the air is always very moist, and this is particularly undesirable for keeping the oil in good condition, the air as it flows into the container is made to pass a long distance over a drying agent (silica gel).

In order to make the necessary connections between the panels there is a horizontal strip with porcelain leads behind every panel, at which the bare wiring lying in the frame terminates and to which the panel is connected. As far as possible the connections on the strips are so arranged that the bare wires do not cross each other. Where crossings were unavoidable, they are concentrated in two of the horizontal strips: these strips are constructed as closed cans in which the connections between the leads on one side (toward the panel) and the other sides (to the frame) may cross each other as much as is necessary, since the can is later filled with a so-called tropics compound, a pitchlike insulation material with a high softening point. The can is then sealed by soldering. Particular attention must be paid to the earth connection with the high frequency used here. A very good electrical contact between the chassis of the panels and the frame is essential. This is obtained by applying the Schoop process to the overlapping edges of panel and frame [2]).

The receiver

On the subject of the receiver we may be brief. It is a superheterodyne apparatus with an intermediate frequency of about 3 Mc/s and a band width of about 120 c/s (twice the maximum modulation frequency of the transmitter, increased by the amount by which the transmitter frequency may vary due to the abovementioned effects). The frequency of the auxiliary oscillator is here also kept constant by means of a long line. The reaction experienced by the long line from the rest of the connections is in this case much less than in the transmitter because of the much smaller powers

[1]) With an anode D.C. voltage of about 800 volts (due to the high temperature of the surroundings the valves work with a decreased load, namely only 60 per cent of the permissible anode voltage, which is 1 200 volts) peak voltages of about 3 200 volts occur at full modulation.

[2]) In this process a firmly adhering layer of metal is deposited by spraying.

(smaller amplifier valves). The output of the receiver is adapted to the impedance of the line to the telephone exchange (normally 600 ohms) by a transformer.

The construction of the receiver is quite analogous to that of the transmitter. With the appurtenant supply apparatus it is assembled in the same kind of case as the transmitter, and the same considerations are valid for the wiring, etc. The heat problem was here much less serious of course, since the receiver consumes only 110 W.

The operation of the installation

The receivers of the two telephone exchanges in radiotelephonic connection must remain in action permanently during business hours, in order that any calls may be announced. The transmitter on the other hand will be preferably switched off after each call on lines which are not too busy. In order to be able to answer a call immediately, the transmitter must be able to be switched on in the shortest possible time. For this reason as far as possible valves with directly heated cathodes are used which are ready for use several seconds after being switched on. Where because of their more favourable properties (steeper slope, less hum, etc.) indirectly heated cathodes were specifically desirable, the cathodes are kept permanently heated. In order not to shorten the life of the cathodes, the anode voltages are only switched on 3 seconds later than the directly heated cathodes, and then at a value 25 per cent lower than the normal, which is only increased to the normal value after an additional 10 seconds. It is thus possible to speak 3 seconds after the transmitter has been switched on, although with a higher interference level for the first 10 seconds, since the lowered anode voltages involve a less favourable signal-noise ratio at the receiving end. The switching on in steps also limits the magnitude of the commutation current surges, so that the fuses introduced at various points can be smaller and the contacts suffer less wear.

The switching on of the transmitter in steps as mentioned takes place by means of a number of relays and bimetallic contacts, all of which are mounted on a panel in the transmitter case (see fig. 6*a*). The first relay, which switches on the receiver and all indirectly heated cathodes, is set in operation from the telephone exchange when the service begins by sending a direct current of 10 mA along the line to the transmitter. When a call is received the telephone operator switches her telemicrophone to this line, and the direct current mentioned is thereby automatically raised to 40 mA by an auxiliary contact. This sets in action a second relay in the transmitter which provides for the successive switching on of the directly heated cathodes, the diminished anode voltages and the normal anode voltages.

In a telephone network a signal with a frequency of 50 c/s or lower is usually used for calling, since the bells of ordinary telephone apparatus are made for this. According to the above description, however, the modulator of the transmitter is not suited to such low frequencies (the frequency characteristic falls off sharply below 300 c/s). Therefore at the telephone exchange at the beginning of the line to the transmitter a separate call apparatus is installed (*fig. 8*). This contains an auxiliary oscillator which is set in action by a relay reacting to 50 c/s, and which gives a signal on the line of 2 700 c/s. At the receiving end this signal, *via* a sharp cut-off filter (also included in the call apparatus), reaches a relay arrangement which now provides the ordinary call signal for the apparatus of the operator. At the same time the current source, switches and milliammeter for the direct currents with which the whole radio installation is operated are included in the call apparatus.

In conclusion it may be mentioned that the transmitter case contains a number of measuring instruments for testing the action of the whole and for detecting any defects. These instruments may be seen in fig. 6*a* on the middle panel; the upper row indicates the current in the different stages of the transmitter, to the left and right below the mains voltage supply and the heating current of the output stage are read off, while the centre meter can be connected at will to a number of

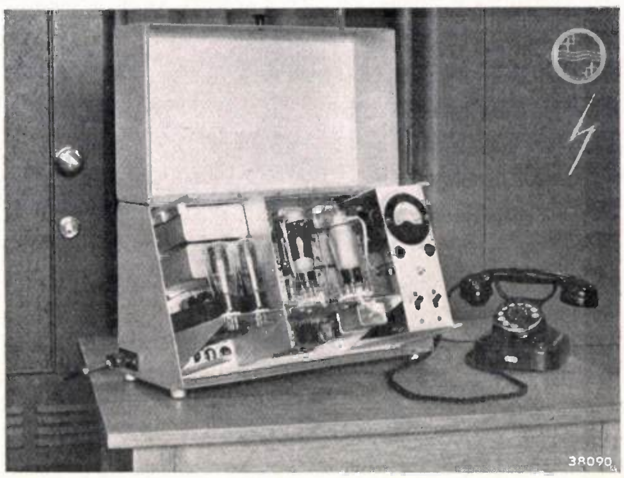


Fig. 8. The auxiliary apparatus in the telephone exchange. In this apparatus are the arrangements for switching on the installation from a distance and for calling the station.

points in the connections by means of the row of contacts to be seen below it. In the upper panel, see fig. 6*a*, a cathode ray tube is installed, to whose deflection plates the low-frequency voltage of the modulator and the high-frequency voltage of the aerial (taken from the aerial supply line led through the top of the case) are applied. On the screen of the tube the modulation characteristic [3]) is obtained in this way, which gives visual evidence of the depth of modulation and any deformation which may be present.

[3]) See H. van Suchtelen, Philips techn. Rev. 3, 248, 1938.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabriken, Eindhoven (Holland), Kastanjelaan.

1520: Th. J. Weijers: Recente ontwikkelingen betreffende frequentiemodulatie (Recent developments in frequency modulation) (T. Ned. Rad. Genoot. **7**, 315-365, Oct. 1940).

A method was proposed by Armstrong in 1936 of obtaining radio reception with less interference due to noise by means of frequency modulation than is possible with amplitude modulation. In this lecture before the Netherlands Radio Association a theoretical discussion was given of the frequency modulation which in its main features follows the line developed by Carson and Fry in 1937, but which is somewhat simpler in various details. It is found that by making use of frequency modulation it is possible to improve the quality of the reproduction much more than is otherwise possible without disadvantage to the selectivity. In order to achieve these favourable results it is necessary to use frequency variations which are large with respect to the highest modulation frequency which is to be transmitted. It is, moreover, necessary to provide the receiver with an amplitude limiter.

1521: F. A. Kröger: ZnS, CdS, MnS en mengkristallen in het ternaire systeem ZnS-CdS-MnS. (ZnS, CdS, MnS and mixed crystals in the ternary system ZnS-CdS-MnS). (Chem. Wbl. **37**, 590-596, Nov. 1940.)

In this lecture given before the sections for Colloid Chemistry and Physical Chemistry of the Netherlands Chemical Society, the appearance of mixed crystals in the ternary system of ZnS-CdS-MnS was discussed on the basis of the phase theory. These investigations were carried out in connection with the fluorescent properties which such substances possess.

1522*: K. de Boer: Stereofonische geluidsweergave (Stereophonic sound reproduction) (diss. Delft Dec. 1940).

In this dissertation a survey is given of the investigations which have been carried out on the subject of three-dimensional hearing and three-dimensional reproduction of sound. In addition to the work of the author himself the results obtained by others are also discussed. The reader is referred to the articles which have already appeared on this subject in this periodical **4**, 316, 1939, **5**, 107 and 182, 1940.

1523: A. Bouwers: Een toestel voor de localiseering van projectielen (An apparatus for the localization of projectiles) (Ned. T. Geneesk. **84**, 4665-4667, Nov. 1940).

For a detailed description of this "bullet finder" the reader is referred to: Philips techn. Rev. **5**, 309, Nov. 1940.

1524: J. A. M. van Liempt and J. A. de Vriend: The fusing time of fuses. III (Z. Phys. **117**, 18-19, Dec. 1940) (Original in German).

Earlier investigations by the authors on the fusing time of thin fuses as dependent upon the short-circuit current are continued for the cases of different metals and alloys. The measurements were carried out with direct current (according to a method already described in this periodical **4**, 118, 1939) with wires of different shaped cross sections. As was to be expected theoretically the shape of the cross section of the wire has no effect on the fusing time.

1525: K. F. Niessen: Electrical field strength as a function of the energy consumed by the aerial II (Physica **7**, 897-908, Dec. 1940) (Original in German).

In continuation of **1510**, the vertical electrical field is calculated which is excited by a half wave length aerial at a point above a flat dry earth, for which a dielectric constant of 4 and a conductivity of 10^{-15} may be assumed. As in **1510** the calculations begin with the simple case where a mathematical dipole is situated at a distance of a half wave length above the earth. As in the earlier case a wave length of 30 m is used; the essential difference from the first publication consists in the fact that now the absolute value of the complex index of refraction is not large compared with unity, and the argument becomes practically equal to 45°, since the displacement current is large compared with the conduction current. The article ends with a comparison of the results found here with the values which would be obtained with the reflection formula. The striking point is not so much the improvement which the formula of Sommerfeld gives compared with the reflection formula, but particularly the fact that the field has now been successfully calculated from the total energy applied to the aerial, in which of course the energy is also included which is later absorbed by the earth.

1526: J. A. M. van Liempt and J. A. de Vriend: Pupil measurements with monochromatic light. (Physica **7**, 961-969, Dec. 1940) (Original in German).

The size of the pupil with monochromatic light has been measured as a function of the wave length and brightness with the help of Broca's pupilmeter constructed by Moss in the form of spectacles. With increasing wave length the pupil becomes larger, with a sudden increase at 5 700 to 5 900 Å. With mixed light there is no simple additivity in the size of the pupil. The summation law, as given by v. Kreveld for the density of photographic material under mixed light, and by Bouma for the subjective brightness of a mixed radiation, probably holds for the contraction of the pupil.

1528: J. D. Fast: Diffusie van gassen door metalen (Diffusion of gases through metals). (Chem. Wbl. **38**, 2-8 and 18-23, Jan. 1941).

A survey is given of the different mechanisms of diffusion which may be present in metals. In those cases where anything is known about the mechanism of the diffusion of the gases through metals, gas atoms (or ions) are found to diffuse through the interstices of the metal lattice. Upon the occurrence of a passage of gases through metal walls, however, it is not exclusively a question of this diffusion process, but also of two processes at the entrance surface and two at the exit surface. Upon transmission of gases through metals therefore there are different processes, each of which possesses its own activation energy. For different combinations of gas and metal and upon the use of not too thick walls, five different cases can also be realized, in each of which a different process determines the velocity of transmission. With the help of potential curves the different possibilities are represented diagrammatically. The writer then examines which of these cases occur in practice. Upon the diffusion of hydrogen through iron the greatest activation energy is necessary at the entrance surface for splitting the molecule into atoms. In the diffusion of hydrogen through palladium and of oxygen through nickel, however, the greatest energy is required at the exit surface for the recombination of the atoms to molecules, *i.e.* for exactly the opposite process. When hydrogen penetrates through copper walls it is a case in which the greatest activation energy is necessary for the diffusion of the hydrogen atoms (or ions) inside the metal.

VOL. 6. No. 5, p. 129-160 MAY 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

THE CAUSES OF VOLTAGE AND CURRENT FLUCTUATIONS

by C. J. BAKKER. 621.396.645.3: 621.396.822

There are in principle only two causes for the occurrence of spontaneous current and voltage fluctuations in amplifiers: the thermal agitation of the electric charge in conductors and the motion of free electrons in amplifier valves. These "sources of noise" in amplifiers, which have been repeatedly discussed in this periodical, are here dealt with once again on the basis of new information which has lately been gained. Special emphasis is laid on the significance of cosmic phenomena in the noise on aerials and on the disturbing effects of the induced currents which are excited by the fluctuations in the emission of the cathode on the control grid of an amplifier valve.

The weak spontaneous current and voltage fluctuations which occur in every electrical circuit have repeatedly been discussed in this periodical [1]). These current and voltage fluctuations may give rise to undesired phenomena when they occur in an apparatus or a part of an apparatus in which extremely small currents and voltages are used, such as a microphone, a photocell or the aerial and the input circuit of a radio receiving set. In the last case the effect of the spontaneous current and voltage fluctuations is universally familiar; they lead to a continuous sound which is usually called the noise of the receiver.

As causes of the noise in receivers, in the articles referred to above, the resistances and the amplifier valves present in the circuit were mentioned. During the three years since the publication of these articles new information has been gained about the spontaneous current and voltage fluctuations in such circuit elements, which information is found to be of special importance in the case of reception on short waves. Since the combatting of the characteristic noise of the set is found to form one of the chief problems in the construction of short wave receivers, we shall again discuss the voltage and current fluctuations in resistances and electronic valves on the basis of the new information, while the results produced by the fluctuations on the performance of the receiving sets will be dealt with in a following article in this periodical.

[1]) M. Ziegler. The causes of noise in amplifiers, Philips techn. Rev. **2**, 136, 1937; Noise in amplifiers contributed by the valves, Philips techn. Rev. **2**, 329, 1937; Noise in receiving sets, Philips techn. Rev. **3**, 189, 1938.

Voltage and current fluctuations in resistances

The voltage and current fluctuations which occur in resistances may be considered as a thermal phenomenon: the more or less mobile electrical charge which is situated in the interior of a resistance is not only brought into motion by an applied voltage, but it also takes part in the thermal molecular motion, just as in the case of microscopically small particles suspended in a liquid (Brownian movement). This motion of the charge is manifested as an irregularly varying current of which it may be proved that in intensity and character it is entirely determined by the magnitude of the resistance and the temperature, while the nature of the resistance (for instance, whether carbon or wire resistance) is without effect [2]).

If the spontaneous current fluctuations are resolved into components with different frequencies, a continuous spectrum is found in which all frequencies occur in equal intensity, *i.e.* each frequency band Δf furnishes the same contribution to the effective current, independent of the frequency f itself. The mean square value of this current contribution is given by the formula

$$\overline{i_R^2} = \frac{4\,kT}{R}\Delta f, \quad \ldots \ldots \quad (1)$$

[2]) The thermodynamic considerations amount to the following: Two resistances A and B which have the same temperature are considered to be connected in parallel: the voltage fluctuations generated by A must then heat resistance B by the same amount as the fluctuations generated by B heat A. Otherwise there would be a temperature difference which would be contrary to the fundamental laws of thermodynamics. See on this subject also the first article referred to in footnote [1]).

where $k = 1.38 \times 10^{-23}$ watt sec/degree represents the Boltzmann constant, T the absolute temperature (°K), R the resistance in ohms and f the frequency in c/s. The formula determines the fluctuation current in amperes which occurs when the resistance is short circuited. The fluctuation voltage v_R between the ends of an open resistance is often desired. This is connected with i_R by $v_R = R i_R$, and therefore

$$\overline{v_R^2} = 4\, kTR\, \Delta f \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

In order to give some idea of the order of magnitude of the fluctuations let us consider a resistance of 10^5 ohms. At room temperature and with a frequency band of 10^4 c/s, equation (1) gives an effective current of

$$\sqrt{\overline{i_R^2}} = 4 \cdot 10^{-11}\ \mathrm{A},$$

while from equation (1a) the following voltage fluctuations result:

$$\sqrt{\overline{v_R^2}} = 4\ \mu V.$$

Up to this point the fluctuation phenomena were also discussed in the article referred to above on the causes of noise in amplifiers. If we now apply the results to actual connections, various questions arise whose answers are not immediately obvious. The first question is the following. One of the resistances which may cause considerable disturbance by their voltage fluctuations is the resistance of the tuned oscillator circuit which is connected between control grid and cathode of the amplifier valve (see *fig. 1*). When we wish to calculate the voltage fluctuations on the grid, what value of the resistance must we choose, the resistance of the coil (r), which may be considered as the actual source of the fluctuations, or the resonance resistance R_{LC} of the whole LC circuit (L/Cr) which is connected between control grid and cathode?

The answer is that this must be a matter of indifference when the calculation is correctly performed. Since the voltage fluctuations over a resistance of a given size are independent of the nature of the resistance, the voltage fluctuations over the oscillator circuit are determined by the resonance resistance R_{LC} in the neighbourhood of the resonance frequency. The mean square of the fluctuation voltage must therefore have the value

$$\overline{V_C^2} = 4\, kTR_{LC}\, \Delta f \quad . \quad . \quad . \quad . \quad . \quad (2)$$

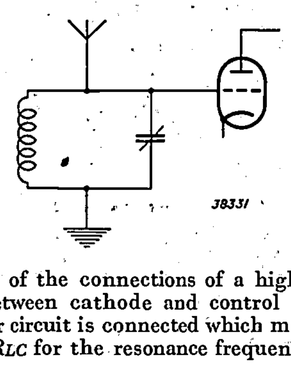


Fig. 1. Diagram of the connections of a high-frequency amplifier stage. Between cathode and control grid a high-frequency oscillator circuit is connected which may be considered as a resistance R_{LC} for the resonance frequency.

The first mentioned argument is, however, also correct; the source of the fluctuations is exclusively the series resistance r of the coil, see *fig. 2* (other sources of damping, such as that of the condenser and the input damping of the valve will here be neglected), and when we assume a voltage fluctuation of

$$\overline{v^2} = 4\, kT\, r\, \Delta f$$

in series with the resistance r, we must also obtain the correct result.

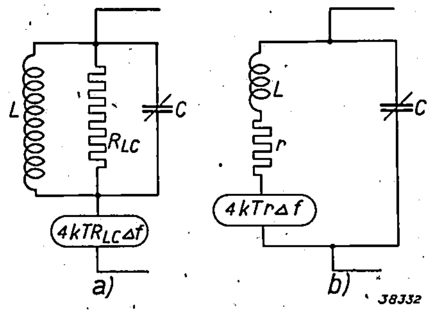


Fig. 2. *a*) Voltage fluctuations in series with the resistance R_{LC} of the oscillator circuit. *b*) Voltage fluctuations in series with the resistance r of the coil. If the damping of the oscillator circuit is caused only by the resistance r of the coil, the two diagrams are equivalent to each other.

The fact that the two lines of reasoning are actually equivalent is confirmed by simply calculating the voltage on the oscillator circuit, beginning with the fluctuation voltage over the resistance r. For this purpose we consider a component of the fluctuations having a given angular frequency, ω. The voltage fluctuations v then cause a current through the oscillator circuit which is given by

$$i = \frac{v}{r + j\left(\omega L - \dfrac{1}{\omega C}\right)}.$$

This current causes between control grid and cathode of the amplifier valve an A.C. voltage:

$$V_C = \frac{i}{j\omega C} = \frac{v}{(1 - \omega^2 LC) + j\omega r C}$$

and by taking the mean square over the time the following result is obtained:

$$\overline{V_C^2} = 4\,kT\,\Delta f \frac{r}{(1-\omega^2LC)^2+\omega^2r^2C^2} \quad (3)$$

In the special case of the resonance frequency, $\omega^2LC = 1$, and therefore

$$\overline{V_C^2} = 4\,kT\,\Delta f \cdot \left(\frac{L}{Cr}\right).$$

The expression in parenthesis is the resonance resistance R_{LC} of the oscillator circuit, so that equation (2) is confirmed.

In a frequency region outside the resonance frequency the following formula is valid for the impedance in parallel with the oscillator circuit:

$$Z_{LC} = \frac{r + j\omega L\left(1-\omega^2LC-\dfrac{r^2C}{L}\right)}{(1-\omega^2LC)^2+\omega^2r^2C^2} = X + jY.$$

The real part of this corresponds to the fraction in equation (3). From this it follows that the voltage fluctuations on an oscillator circuit for any given frequency can be described by the general formula:

$$\overline{V_C^2} = 4\,kT\,X\,\Delta f, \quad . \; . \; . \; . \; . \quad (4)$$

where X is the real part of the circuit impedance.

In order to be able to use equation (4) the frequency interval Δf must be chosen so small that within this interval the circuit impedance does not change to any extent. If one is concerned with a finite frequency interval, $X\Delta f$ must be replaced by $\int X\mathrm{d}f$. The exact expression for the mean square of the fluctuation voltage then becomes:

$$\overline{V_C^2} = 4kT\int_{f_1}^{f_2} X\mathrm{d}f = 4kT\frac{r}{2\pi}\int_{\omega_1}^{\omega_2}\frac{\mathrm{d}\omega}{(1-\omega^2LC)^2+\omega^2r^2C^2}.$$

By the introduction of the resonance frequency $1/\sqrt{LC} = \omega_0$ and of the symbols $\omega/\omega_0 = x$ and $\omega^2r^2C^2 = a^2$ this formula can be written:

$$\overline{V_C^2} = \frac{4kT}{C}\,\frac{a}{2\pi}\int_{x_1}^{x_2}\frac{\mathrm{d}x}{(1-x^2)^2+a^2x^2} \quad . \; . \; . \quad (4a)$$

The integration is somewhat complicated, but does not offer any fundamental difficulties. The general integral is:

$$\frac{4kT}{C}\cdot\frac{a}{2\pi}\left[\frac{1}{2a}\,\mathrm{arctg}\,\frac{ax}{1-x^2} + \frac{1}{4\sqrt{4-a^2}}\ln\frac{x^2+x\sqrt{4-a^2}+1}{x^2-x\sqrt{4-a^2}+1}\right]_{x_1}^{x_2}.$$

We shall now consider two special cases more closely. Firstly the case, which is interesting from a physical point of view, where the integral extends over all frequencies ($x_1 = 0$, $x_2 = \infty$), secondly the case which is more important technically where only a certain region on either side of the resonance frequency furnishes a contribution.

In the first case the inverse tangent occurring within the bracket is found to possess a value π, while the second term between the brackets furnishes no contribution at the limits 0 and ∞. The result therefore is that

$$\overline{V_C^2} = \frac{4kT}{C}\,\frac{a}{2\pi}\,\frac{1}{2a}\cdot\pi = \frac{kT}{C} \quad . \; . \; . \; . \; . \quad (4b)$$

The total fluctuation voltage on the condenser is thus independent of the self-induction and the resistance in the oscillator circuit.

The result could indeed have been derived directly by a thermodynamic method. According to the law of equipartition, the following holds for the potential energy of the charged condenser:

$$\tfrac{1}{2}C\,\overline{V_C^2} = \tfrac{1}{2}kT,$$

from which (4b) follows directly.

We shall now pass on to the case where only a limited frequency region on either side of the resonance frequency furnishes a contribution to the fluctuation voltage. The width of this region is in general determined by the selectivety of the following stages, and may for example be chosen such that the resonance resistance has fallen to one half at the limits of the region.

From equation (4a) it may be seen that for the limiting frequencies x_1 and x_2 the absolute value of $1-x^2$ is then equal to ax. The argument of the inverse tangent in the general integral then has the value -1 at the lower limit and the value $+1$ at the upper limit. The inverse tangent thus covers a range of angles of 90° or $\pi/2$, namely from $-45°$ to $+45°$. The second term between the brackets, in the important case of a sharp resonance curve ($a \ll 1$), is again found to furnish no contribution. The result is therefore:

$$\overline{V_C^2} = \frac{4kT}{C}\cdot\frac{a}{2\pi}\cdot\frac{1}{2a}\cdot\frac{\pi}{2} = \frac{kT}{2C} \quad . \; . \; . \; . \quad (4c)$$

The fluctuation energy in the frequency region chosen is therefore one half of the total fluctuation energy.

For practical applications it is advantageous to write equation (4c) in a form which more nearly resembles equation (4), namely:

$$\overline{V_C^2} = 4kT\,\overline{X}\,\Delta f, \quad . \; . \; . \; . \; . \; . \; . \quad (4d)$$

where $\overline{X}$ is the average value of the real part of the circuit impedance in the frequency region in question. By the use of the integral given in the foregoing this average value can easily be calculated and one obtains:

$$\overline{X} = \frac{\pi}{4}\left(\frac{L}{Cr}\right) = 0{,}78\,R_{LC},$$

while for the frequency region at whose limits the real part of the circuit impedance has fallen to one half, a width of

$$\Delta f = \frac{1}{2\pi}\,\frac{r}{L}$$

is calculated. By substituting these values of $\overline{X}$ and Δf in equation (4d), formula (4c) should be obtained. One actually finds that:

$$\overline{V_C^2} = 4kT\left(\frac{\pi}{4}\,\frac{L}{Cr}\right)\cdot\left(\frac{1}{2\pi}\,\frac{r}{L}\right) = \frac{kT}{2C}.$$

If the sources of damping do not lie in the self-induction coil, but in the condenser, fundamentally the same results are found. For dampings which are caused by the amplifier valve, however, different results may be expected, about which some remarks will be made later on in this article.

Current fluctuations in the aerial

An oscillator circuit of special interest to us is the aerial. It has a definite impedance depending upon the frequency, the real part of which is usually called the radiation resistance of the aerial for the frequency in question.

This radiation resistance, according to the general law (1a) will also have to exhibit certain thermal fluctuations. This involves the question as to what its temperature is. It is clear that the temperature of the aerial wire has nothing to do with the fluctuations in question. The radiation resistance R_{ant} of the aerial means physically that when an A.C. voltage V_{ant} is applied to the aerial it dissipates an energy of:

$$P = \frac{\overline{V_{ant}^2}}{R_{ant}}$$

which is radiated into space. Conversely, the thermal voltage fluctuations must have their source in the fact that space is filled with thermal radiation (heat radiation) which is taken up by the aerial. If it is assumed that there is in space an equilibrium between the radiation and matter at a given temperature T, then the intensity and special distribution of the radiation is known. It is therefore possible to calculate the voltage fluctuations on the aerial caused by the radiation, and one finds, as was to be expected, that between the radiation resistance and the voltage fluctuations of an aerial there is the general relation:

$$\overline{V_{ant}^2} = 4\,kT\,R_{ant}\,\Delta f$$

How high is the effective temperature T of space for heat radiation of radio frequencies? The voltage fluctuations on an aerial make it possible to give an experimental answer to this question. Experiments in this direction carried out by Jansky and Reber [3]) have led to very remarkable results. It is found that the radiation in the region of the high radio frequencies (metre waves) is mainly of cosmic origin and appears to come from the milky way. Instead of a temperature of several degrees absolute, as might be expected for the universe, a temperature of about 10 000 °K is found.

On the basis of theoretical investigations [3]) it may be accepted as practically certain that the radiation which causes the noise of the aerial has its origin in interstellar matter which fills the milky way system with a concentration by weight corresponding to approximately one hydrogen atom per cm³. The thickness of the layer is about 50 000 light years. Under the action of starlight this matter is ionized for the main part into electrons and positively charged particles. The irregular motion of the electrons, as a result of their mutual repulsion and their attraction of the positive particles, leads to a radiation which is intense enough to make the observed phenomena understandable in principle. A complete explanation will, however, only be possible when more data are available about interstellar matter.

Voltage and current fluctuations in amplifier valves

The currents in an amplifier valve have their source in the thermal emission of a heated cathode. If this process is considered fluctuation phenomena are immediately expected because of the fact that the negative charge emitted consists of electrons, *i.e.* particles, and is therefore emitted in definite multiples of $e = 1.6 \times 10^{-19}$ coulombs.

The number of electrons which is emitted by the surface of the cathode in a time interval of a given length will not always be exactly the same, but will exhibit accidental variations around a definite average value, just as, when it is raining, the number of drops which strike a definite spot on the ground is not always exactly the same for every second. When it is assumed that all the electrons emitted reach the anode the variations experienced by the anode current due to these fluctuations can easily be calculated. For the performance of this calculation we may again refer to the article cited in footnote [1]) on the causes of noise in amplifiers. The calculation leads to the result that the current variations, like the thermal current fluctuations in a resistance, cover a continuous spectrum with constant amplitude. The mean square of the current fluctuations i_k is given by the formula:

$$\overline{i_k^2} = 2\,e\,I_k\,\Delta f, \qquad (5)$$

where I_k represents the average emission current of the cathode [4]).

With the practically occurring values of $I_k = 30$ mA, $\Delta f = 10\,000$ c/s, for example, one obtains

[3]) K. G. Jansky, A note on the source of interstellar interference, Proc. Inst. Rad. Eng. **23**, 1158, 1935. G. Reber, Cosmic static, Proc. Inst. Rad. Eng. **28**, 68, 1940; Astrophys. J. **91**, 621, 1940. For theoretical considerations of this radiation see: H. A. Kramers, Theory of the continuous X-ray spectrum, Phil. Mag. **46**, 836, 1932; A. S. Eddington, Diffuse matter in interstellar space, Proc. Roy. Soc. **111**, 424, 1926. J. A. Gaunt, Continuous absorption, Trans. Roy. Soc. **229**, 163, 1930. L. C. Henyey and P. C. Keenan, Interstellar radiation from free electrons and hydrogen atoms, Astrophys. J. **91**, 625, 1940.

[4]) This result is valid only for frequencies at which the time of oscillation is still several times greater than the transit time of the electrons between cathode and anode. On the subject of the current fluctuations at still higher frequencies see S. Ballantine, J. Franklin Inst. **206**, 159, 1928.

$\sqrt{\overline{i_k^2}} = 10^{-8}$ A. This current fluctuation caused by the corpuscular nature of the electric charge, is usually called the shot effect of the amplifier valves.

By measurements with a diode the formula for the shot effect may be tested. If the anode voltage is chosen sufficiently high, the condition is indeed fulfilled that each electron which leaves the cathode reaches the anode. The diode is then said to be in the saturated state, since upon further increase of the anode voltage the anode current no longer increases. In this state of saturation equation (5) is found to be exactly confirmed. The constant e appearing in the equation can be so accurately derived from the shot effect that this may be considered as one of the best methods of determining the charge on the electron.

Anode current fluctuations in amplifier valves

In the case of the customary amplifier valves (triodes and pentodes) the anode and the grid voltages are chosen so low that there can be no question of saturation. Otherwise indeed it would be impossible to regulate the anode current by variation of the control grid voltage. It is now found that the current fluctuations in the normal state of working of the amplifier valves are considerably smaller than would be expected according to equation (5). Qualitatively, this is easily understood. The electrons which are situated between cathode and control grid while the valve is in action (space charge) exert a repulsive effect on all electrons which are on the point of leaving the cathode. Because of this, part of the electrons emitted are driven back to the cathode, so that the anode current is much smaller than the emission of the cathode. When the cathode emission now fluctuates so that at a certain moment a larger number of electrons are emitted than the average, the space charge also becomes stronger and sends a larger percentage of the emitted electrons back to the cathode. The space charge thus smooths the current fluctuations. Experimental data on this smoothing in amplifier valves in practical use have already been given in this periodical [5]). The smoothing is expressed by a "noise factor" F, which indicates the ratio between the current fluctuations actually occurring and the current fluctuations which would be expected according to the shot effect. Expressed in a formula

$$F^2 = \frac{\overline{i_k^2}}{2\, e\, I_k\, \Delta f} \qquad (6)$$

[5]) See the second article mentioned in footnote [1]).

The value of F^2 determined experimentally according to this formula is always less than unity in the case of ordinary amplifier valves, and in many cases it amounts to only a few per cent.

We shall now go somewhat more deeply into the theoretical basis of the smoothing of the current fluctuations by the charge.

Theory of the suppression of the anode current fluctuations by the space charge.

A simple consideration, which makes it understandable that in the presence of a space charge the current fluctuations disappear in the first approximation, is possible on the basis of the theory of the space charge given by Langmuir [6]). Langmuir calculated for a diode with a plane cathode and a plane anode, the anode current per unit surface which may flow with a given anode voltage, when it is not the emission of the cathode but the repulsive action of the space charge which limits the current. The result is the well known equation:

$$I_k = \frac{1}{18\,\pi}\,\frac{m}{e}\,\frac{\left(\frac{2e}{m}\,V_a\right)^{3/2}}{d^2}, \qquad (7)$$

where e is again the charge and m the mass of the electron. V_a is the anode voltage and d is the distance between cathode plate and anode plate.

The important feature of this formula is that the saturation current I_s of the cathode does not occur in the result. The result is of course only valid when I_s is much greater than I_k.

Since the emission does not occur in the result it is also immediately clear that accidental fluctuations of the emission are unable to affect the magnitude of the current passing. If Langmuir's formula is correct, therefore, a complete suppression of the current fluctuations could be expected. As already mentioned the experimentally found suppression is indeed very strong. A certain residual effect remains, however, and indicates that the conception of the limitation of the current by the space charge as given here is too general and does not exactly correspond to the actual facts.

By means of a simple improvement in the calculations it is possible to explain the residual effect and to calculate it. In the derivation of equation (7) an error of neglect has been made by assuming that the electrons leave the cathode without initial velocity. Because of the fact that the electrons actually have a certain initial velocity, the space charge in the neighbourhood of the cathode is

[6]) I. Langmuir, Phys. Rev. 2, 450, 1913; 21, 419, 1923.

slightly lower, so that the limit of the current prescribed by the space charge will be slightly greater.

The way in which the space charge limits the current is shown graphically in *fig. 3a*. Between cathode and anode a potential minimum occurs due

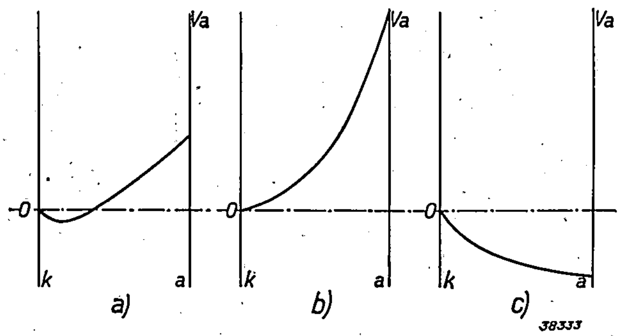


Fig. 3. Variation of the potential between cathode k and anode a of a diode, a) for the ordinary state of operation. b) for very high anode voltages, c) for very low (negative) anode voltages. In the ordinary state of operation the potential between cathode and anode exhibits a minimum.

to the space charge. Between the cathode and this minimum there is a retarding field for the electrons, which can only be overcome by those electrons whose initial velocity lies above a certain value. The other electrons return to the cathode. Furthermore, it may be seen from the figure that this minimum disappears for very high anode voltages (3b), as well as for sufficiently negative anode voltages (3c). In these limiting cases the regulating action of the space charge actually disappears, and the unimpaired shot effect is again obtained.

The correction of the initial velocity is easily added to Langmuir's calculation. One then finds for the first approximation of the current:

$$I_k = \underset{(u=0)}{I_k}\left(1+\frac{3\overline{u}}{\sqrt{\frac{2e}{m}V_a}}\right), \quad . \quad . \quad . \quad (8)$$

where $\overline{u}$ is the average initial velocity of the electrons in a direction perpendicular to the cathode.

This velocity is determined by thermal factors; it corresponds approximately to the velocity which the electrons would possess in a gas at the temperature T_k of the cathode. It is known that the average kinetic energy of an atom or electron in a gas is equal to $^1/_2 kT_k$ per degree of freedom, so that $\overline{u^2}$ is of the order of magnitude of kT_k/m.

Like all thermal phenomena, the initial velocity exhibits certain fluctuations, and this is why the anode current is not absolutely constant. If we consider a certain time interval τ, the average velocity $\overline{u_\tau}$ of the electrons emitted in that time interval will in general differ from $\overline{u}$. If we set $\overline{u_\tau}-\overline{u} = \delta u$, then according to equation (8) the current fluctuations are:

$$\overline{i_k^2} = I_k^2\,\frac{9\,\overline{\delta u^2}}{(2e/m)\,V_a}. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The fluctuations of the initial velocity of the electrons are of the same order of magnitude as the initial velocity itself. If in each time interval τ one electron should be emitted, $\overline{\delta u^2}$, like $\overline{u^2}$, would amount to about kT_k/m. If n_τ is the average number of electrons which is emitted per time interval τ, the fluctuation of the *average velocity of these n_τ electrons* is smaller by a factor n_τ, so that one may write:

$$\overline{\delta u^2} = a\,\frac{kT_k/m}{n_\tau} = a\,\frac{kT_k/m}{I_k\,\tau/e}, \quad . \quad . \quad (10)$$

where a is a numerical factor of the order of magnitude of unity. An exact calculation gives: $a = 2(1-\pi/4) = 0.429$.

The square of the velocity fluctuations of the electrons is thus proportional to the cathode temperature T_k, and inversely proportional to the specific time interval τ. This is exactly the same relation as would be found for the thermal current fluctuations in a resistance if they were considered, not for a given frequency band, but for a given time interval. Conversely, in the formula for the velocity fluctuations we may also pass over from the time interval τ to a frequency interval Δf. A simple harmonic analysis shows that it is only necessary to substitute $2\Delta f$ for $1/\tau$ [7]). For the velocity fluctuations of the emitted electrons one then finds

$$\overline{\delta u^2} = 2\,a\,\frac{kT_k/m}{I_k/e}\,\Delta f$$

and by introducing this result in (9) one obtains the anode current fluctuations:

$$\overline{i_k^2} = I_k^2\,\frac{9}{(2\,e/m)\,V_a}\,2\,a\,\frac{kT_k/m}{I_k/e}\,\Delta f =$$
$$= 9\,a\,kT_k\,\frac{I_k}{V_a}\Delta f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The most striking part of this result is that the properties of the electron (charge and mass) have disappeared entirely from the final formula. The current fluctuations of a diode whose anode current is limited by space charge are apparently a thermal phenomenon. The relation found for them very

[7]) The same transition from time intervals τ to frequency intervals Δf was carried out for the shot effect earlier in this periodical, and the following formulae were obtained:

$$\overline{i_k^2} = \frac{e\,I_k}{\tau} \quad \text{et} \quad \overline{i_k^2} = 2\,e\,I_k\,\Delta f$$

See the first article cited in note [1]).

much resembles the one for the thermal current fluctuations of a resistance (equation 1), since the quotient V_a/I_k may be considered as the internal resistance of the diode. For practical applications it is advantageous to introduce the slope $S = \mathrm{d}I_k/\mathrm{d}V_a$ of the diode instead of the quotient V_a/V_k. Since the anode current varies with the 3/2 power of the anode voltage

$$S = \frac{3}{2}\frac{I_k}{V_a} \text{ and therefore}$$

$$\overline{i_k^2} = 0{,}644 \cdot 4\, kT_k S\, \Delta f \quad . \; . \; . \; . \quad (12)$$

Finally we may again express the fluctuations by the noise factor defined in equation (6). By combining (6) and (11) we then find:

$$F^2 = \frac{9\,\alpha\, kT_k}{2\,e\,V_a}\,.$$

F is thus proportional to $1/\sqrt{V_a}$.

The result is only valid for anode voltages for which the approximations introduced during the calculations are permissible. This is the case only when V_a is not too low, but on the other hand far enough below the saturation value V_s. If the anode voltage lies outside this interval the calculation becomes much more difficult, and can only be carried out numerically [8]).

Fig. 4 gives a graphical idea of the result for a special case. The region in which our approximations are valid is indicated by dotted lines. Outside this region the noise factor increases suddenly; for negative anode voltages and anode voltages greater than V_s the noise factor is equal to unity.

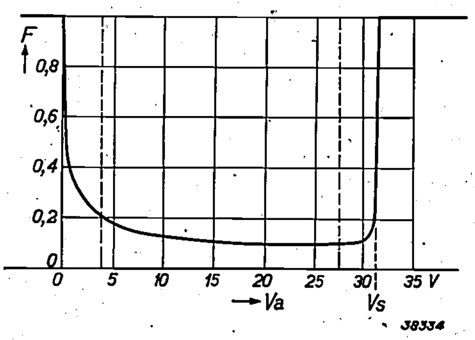


Fig. 4. Noise factor F as a function of the anode voltage. In the sloping part the variation is calculated according to equation (12), while in the neighbourhood of $V_a = 0$ and $V_a = V_s$ numerical calculations have been applied.

The results are well confirmed experimentally. In carrying out measurements on a diode larger values are usually found, it is true, than given by the theory, since electrons are reflected by the anode and disturb the field distribution and thereby cause additional fluctuations. If, however, a triode is used, these difficulties do not occur.

If equations (11) and (12) are applied to a triode the anode voltage must be replaced by the "effective control voltage" V_{eff} in the grid plane. This control voltage may be calculated from the anode current with the help of Langmuir's formula. The quantity S in equation (12) is now defined by $\mathrm{d}I_k/\mathrm{d}V_{\mathrm{eff}}$, *i.e.* the change in the anode current with the control voltage. If the anode voltage remains constant, the change in the control voltage is smaller than the change in potential of the grid wires themselves, since the anode potential also acts in the grid plane. The value of S to be used in equation (12) is therefore greater than the measured slope of the triode, the difference may amount to a factor 2, for instance.

Other fluctuation phenomena in triodes

The spontaneous fluctuations of the current which leaves the cathode are important not only because of their direct effect on the anode circuit, but also because of the fact that the alternating current upon passing the negative control grid wires induces an alternating electrical charge on the control grid. Since in general there is a very high impedance between control grid and cathode, the control grid current which corresponds to the fluctuations of the control grid charge may lead to considerable voltage variations. These variations in turn again cause fluctuations of the anode current. These fluctuations are found in the case of short waves to be able to furnish a very considerable contribution to the noise of receiving sets, and therefore require our attention [9]).

The negative charge Q which is present between the electrodes of an amplifier valve is given by the product of the current I_k and the transit time τ_{ka} of the electrons between cathode and anode. This charge leads to an equally large induced charge which is distributed over the different electrodes; a certain part of it q is therefore also present on the control grid. If the current I_k exhibits certain fluctuations i_k, the space charge $I_k\tau_{ka}$ fluctuates by an amount $i_k\tau_{ka}$, and we may therefore write for the fluctuations of the charge on the control grid:

$$\delta q = \alpha\, i_k\, \tau_{ka}\,,$$

where α is a factor smaller than unity.

[8]) E. Spenke, Wiss. Veröff. Siemens Werke **16**, 19, 1937.

[9]) C. J. Bakker, Physica 8, 23, 1841.

In the case of amplifier valves of ordinary construction it is found that the induced charge on the control grid comes mainly from the electrons which are situated between cathode and control grid. If these electrons alone are taken into account, by substituting, instead of the whole transit time τ_{ka}, the transit time to the control grid τ_{ka}, one then calculates *as a first approximation* for a the value $^1/_3$, and one obtains the relation:

$$\delta q = \frac{1}{3}\, i_k\, \tau_{kg} \quad . \; . \; . \; . \; . \; . \quad (13)$$

We shall now imagine the current fluctuations to be resolved into components of different frequencies, as in the foregoing, and shall consider only those components for which the angular frequency lies in the neighbourhood of a certain value ω. The variation δq then leads to a control grid current of the following magnitude:

$$i_g = \frac{1}{3}\, j\omega\tau_{kg}\, i_k. \quad . \; . \; . \; . \; . \; . \quad (14)$$

and for the mean square of the current fluctuations one thus obtains:

$$\overline{i_g^2} = \overline{i_k^2}\, \frac{1}{9}\, \omega^2\, \tau_{kg}^2.$$

If one now substitutes for $\overline{i_k^2}$ in this equation the value according to equation (12), one obtains:

$$\overline{i_g^2} = \frac{1}{9}\, 0{,}644 \cdot 4\, kT_k\, S\Delta f \cdot \omega^2 \tau_{kg}^2\,. \; . \quad (15)$$

Finally the result may be simplified by making use of the fact that the transit time of the electrons leads not only to the grid current just calculated, but also to a damping in the grid circuit. The explanantion and calculation of this "electron input damping" have already been given in this periodical [10]). The result found is:

$$\frac{1}{R_e} = \frac{1}{20}\, S\omega^2\tau_{kg}^2.$$

[10]) The electron input may be explained as follows. If a grid A.C. voltage of not too high a frequency is applied between the control grid and cathode of an amplifier valve, an alternating charge in the same phase is present on the grid, which is manifested in a capacitative grid current 90° in phase ahead of the grid voltage. At very high frequencies, however, it must be taken into account that the charge on the control grid consists partially of the induced charge of the electrons between control grid and cathode. This induced charge follows the voltage variations with a certain time lag which is determined by the transit time of the electrons. The result is a phase shift of the induced current, which thus no longer remains purely capacitative, but takes on a real component in phase with the applied voltage. The calculation of this component, the electron input damping, is outlined in Philips techn. Rev. 1, 176, 1936.

The value of R_e thus obtained is called the electron input resistance of the valve. By substituting the electron input resistance in equation (15) one finds:

$$\overline{i_g^2} = \frac{20}{9} \cdot 0{,}644 \cdot \frac{4\, kT_k}{R_e}\, \Delta f = 1{,}43\, \frac{4\, kT_k}{R_e}\, \Delta f\,. \quad (16)$$

Thus again a formula which closely resembles the one for the thermal current fluctuations of a resistance. We are, however, of the opinion that the numerical difference (in this case the factor 1.43) is essential, and shows clearly that we are here not concerned with a phenonenon which could have been derived by a thermodynamic method.

The grid current fluctuations are determined by the cathode temperature T_k which is about 4 times as high as room temperature. Together with the factor 1.43, therefore, one finds that the electron input resistance causes current fluctuations at least 5 times as large as a resistance of the same size which is connected externally between cathode and control grid. At very high frequencies the electron input damping $1/R_e$ is often greater than the damping $1/R_{LC}$ of the oscillator circuit in parallel with it. The spontaneous current fluctuations in the grid circuit may then form the most important source of disturbance of the receiving set.

The result, equation (16), was tested by means of experiments with an amplifier valve of the type EF 50. As may be seen in *fig. 5*, the grid current fluctuations $\overline{i_g^2}$ are actually proportional to the electron input damping $1/R_e$; they both vary proportionally with ω^2, as the theory demands. The absolute magnitude of the current fluctuations measured also agrees well with the calculated values (see *table I*).

Table 1

Calculated and measured values of the grid current fluctuations $\overline{i_g^2}$ for the three working states indicated in fig. 5 of the pentode EF 50.

I_a (mA)	$\overline{i_g^2}$ (Amp.^{2})	
	measured	calculated
1.20	$1.9 \cdot 10^{-41}\ \omega^2\, \Delta f$	$1.6 \cdot 10^{-41}\ \omega^2\, \Delta f$
4.15	$4.8 \cdot 10^{-41}\ \omega^2\, \Delta f$	$4.5 \cdot 10^{-41}\ \omega^2\, \Delta f$
9.60	$9.3 \cdot 10^{-41}\ \omega^2\, \Delta f$	$7.4 \cdot 10^{-41}\ \omega^2\, \Delta f$

ω angular frequency, f frequency, both in sec^{-1}.

Current fluctuations in valves with more than one positive electrode

a) Distribution fluctuations

In modern receiving sets the first amplifier valve is almost always a screen grid valve, *i.e.* a valve which contains a screen grid in addition to the control grid. The screen grid, which is placed between control grid and anode, is at a constant positive potential and in this way serves to prevent (undesired) changes in the potential field in the neighbourhood of the control grid caused by changes in the anode voltage.

The electrons which pass from the cathode to the anode must therefore pass the screen grid as well as the control grid. Since the former is also at a positive potential, part of the electrons are captured by the grid wires. As a result of this a shot effect again occurs, *i.e.* even when the number of electrons which moves toward the screen grid per time interval τ is absolutely constant, the number of elec-

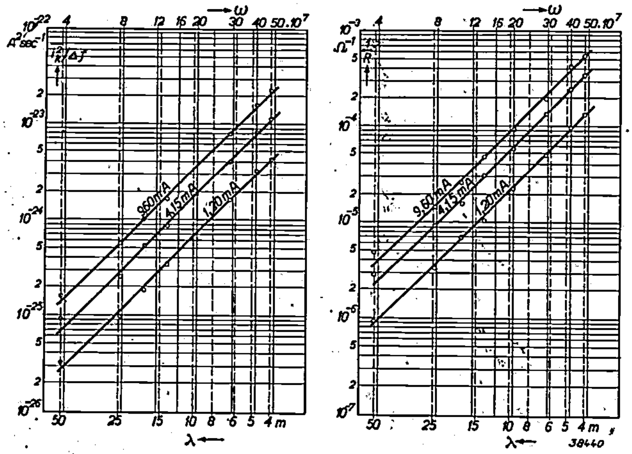


Fig. 5. *a*) Grid current fluctuations $\overline{i_g^2}$,*) *b*) input damping $1/R$ of a pentode EF 50 as a function of the frequency (the figure does not give the electron input damping, but the total input damping). The curves were recorded for three different adjustments of the value with anode currents of 1.20 mA, 4.15 mA and 9.60 mA, respectively. It may be seen that both $\overline{i_g^2}$ and $1/R$ are proportional to ω^2.

trons which impinges on the wires of the grid instead of passing between them will still exhibit certain fluctuations. This causes current fluctuations in the screen grid circuit, and at the same time — with an opposite sign — in the anode circuit. For the size of the fluctuations, the so-called current distribution fluctuations, one may expect equation (5) of the shot effect to be approximately valid:

$$\overline{i_{g2}^2} = 2\, e\, I_{g2}\, \Delta f.$$

This result is only valid when the cathode current exhibits no fluctuations, and when, moreover, the screen grid current I_{g2} is small compared with the total cathode current $I_k = I_{g2} + I_a$. If the cathode current fluctuations given by a noise factor F are included in the calculation, one finds for the total fluctuations in the anode circuit the formula [11]):

$$\overline{i_a^2} = 2\, e\, \frac{I_a}{I_k}\, (I_{g2} + F^2 I_a)\, \Delta f, \quad . \quad . \quad (17)$$

where the first term in parenthesis refers to the distribution fluctuations and the last term to the cathode current fluctuations.

In valves of ordinary construction the screen grid current may amount for instance to 20 per cent of the anode current. The square of the current distribution fluctuation $\overline{i_{g2}^2}$ then also amounts to about 20 per cent of the current fluctuations which the anode current would exhibit in the case of unimpaired shot effect. Now we have seen that the cathode current fluctuations can be weakened by the space charge to less than 10 per cent of their original value. The distribution fluctuations then form the chief source of noise disturbance in the anode circuit.

b) *Fluctuations due to secondary emission*

In order to increase the slope $\mathrm{d}I_a/\mathrm{d}V_g$, of radio valves, more and more use is being made of the phenomenon of secondary emission. *Fig. 6* shows diagrammatically the arrangement of the electrodes in such a valve. The current I_{prim} strikes the auxiliary electrode h which is covered with a secondary emitting substance. The current of secondary electrons is received by the anode a. This current has a value $I_{\text{sec}} = \delta I_{\text{prim}}$ when δ is the average number

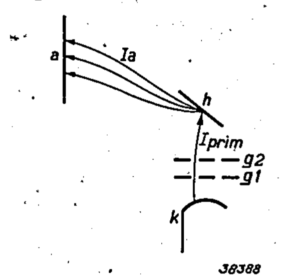


Fig. 6. Electrode system of an amplifier valve with secondary emission.

of secondary electrons liberated per primary electron from the auxiliary electrode.

Like the direct current, the slope is increased by a factor δ. Since the factor δ may amount for instance to 4.5, this is no inconsiderable improvement.

The fluctuations of the anode current may be written in the form:

$$\overline{i_a^2} = \overline{i_{\text{prim}}^2}\cdot \delta^2 + \overline{i_{\text{sec}}^2}, \quad . \quad . \quad . \quad . \quad (18)$$

The first term of this equation indicates that the fluctuations of the primary current are increased by a factor δ. The second term, $\overline{i_{\text{sec}}^2}$, is due to the fact

*) In the figure by error $\overline{i_k^2}$ has been used.

11) C. J. Bakker, Physica 5, 581, 1938.

that a primary electron which arrives on the auxiliary electrode does not always free the same number of secondary electrons, but that in this phenomenon also certain fluctuations occur.

The number of secondary electrons per primary electron may be described by a statistical distribution around an average value. Let $\beta_0, \beta_1 \dots \beta_m \dots$ be the chance that 0,1, ... m ... secondary electrons are freed by a primary electron ($\sum_m \beta_m = 1$), then $\delta = \sum_m \beta_m m$.

It is obvious that a primary electron which frees m secondary electrons will lead to a current fluctuation proportional to $m-\delta$. The absolute value of this current fluctuation i_m is given by an expression which closely resembles the familiar formula for the shot effect, namely:

$$\overline{i_m^2} = 2\, e\, I_{\text{prim}}\, \beta_m\, (m-\delta)^2\, \Delta f.$$

The total contribution to the fluctuation $\overline{i_{\text{sec}}^2}$ is obtained by a summation of this result over all values of m. If, further, we write for the fluctuation of the primary current,

$$\overline{i_{\text{prim}}^2} = 2\, e\, I_{\text{prim}}\, F_{\text{prim}}^2\, \Delta f,$$

we obtain

$$i_a^2 = 2\, e\, I_{\text{prim}}\, [F_{\text{prim}}^2\, \delta^2 + \sum_m \beta_m\, (m-\delta)^2]\, \Delta f\ . \quad (19)$$

In order to apply this formula, for a given material and a given velocity of the primary electrons, one would have to know the probabilities β_m for each value of m. An experimental determination of these probabilities is impossible at present. From measurements on actual valves it is found that the second term of equation (19) may be of the same order of magnitude as the first term, so that the extra current fluctuations caused by the secondary emission furnish no unimportant contribution to the total fluctuations of the anode current.

The last relation is very much simplified if the primary current exhibits the true shot effect [12]), so that $F_{\text{prim}} = 1$. Then according to equation (19):

$$i_a^2 = 2\, e\, I_{\text{prim}}\, (\sum_m \beta_m\, m^2)\, \Delta f, \quad . \ . \ . \quad (20)$$

an equation which can also be directly understood by considering that every term of the sum in parenthesis represents an anode current $I_{\text{prim}} \cdot \beta_m \cdot m$, which consists in each case of portions of m electrons, so that it will furnish a contribution to the shot effect, corresponding to an elementary charge $m.e$.

[12]) M. Ziegler, Physica 3, 1, 1936.

NEW PRINCIPLES OF CONSTRUCTION FOR THE ELECTRO-ACOUSTIC INSTALLATION OF STUDIOS

621.396.712.3 : 534 : 86

by F. DE FREMERY and J. W. G. WENKE.

The studio building which was put into use as an annex of the old A.V.R.O.-studio building in 1940 possesses an installation which differs from the customary one in various important features. The contact between the performing artists and the technical personnel has been made closer than was previously the case. The amplifiers occurring in the installation are not fed from a central battery, but each one is fed from its own built-in supply apparatus. By the standardization of levels and impedances of the connection lines, the number of different types of amplifiers could be reduced to three. Inputs and outputs of the amplifiers can be connected by means of cross connection panels to the different microphones, regulators, lines, etc., whereby with a relatively small number of amplifiers all the possibilities of connection occurring during use can be realized.

A description has already been given in this periodical [1]) of the large studio building of the A.V.R.O. which was inaugurated in 1936. After only a few years, however, the building was already found inadequate for the ever increasing activities connected with broadcasting, so that a start was made on the construction of a new annex studio building. This building which contains two concert studios, and which was finished last year, deviates considerably in arrangement from the old building. In particular a number of important new principles have been employed in the electro-acoustic system, which, like that in the older building, was installed by N.S.F.-Philips. The most important points will be discussed in the following.

The connection of the control tables

There is a certain two-sidedness in the activities of a broadcasting studio: at least in the main programmes the object is the achievement of artistic performances, while on the other hand purely technical questions are always in the foreground. If, for example, we consider the broadcasting of a concert where the sound of different parts of the orchestra and possible soloists is taken up by different microphones, the different microphone contributions must then be regulated separately and mixed in the correct proportions, and at the same time the dynamics of the sound must be adapted to the difficulties of the broadcast, *i.e.* the amplification of the whole must continually be regulated during the performance in order to bring out the pianissimos sufficiently strongly above the natural interference level (noise) and in order not to exceed the maximum permissible depth of modulation in the fortissimos [2]). While the artistic control is in the hands of the conductor, there is a very real danger that his intentions will be nullified by these technical processes, for instance, that the balance of orchestra and soloists will be disturbed or dynamic finesses regulated out of existence.

In order to avoid this, in addition of course to a good training of the mixing and regulating technicians, a visual contact between the latter and the conductor is also desired. In the old A.V.R.O.-building the necessity was met by glass cabins in the studios, in which the person is seated who mixes the contributions of the different microphones in the studio in question. The regulation of the whole broadcast then takes place at a control table, which is placed in a separate room (*fig. 1*), and on which, in addition to the control instruments (modulation meter and monitor loud speaker), the regulating

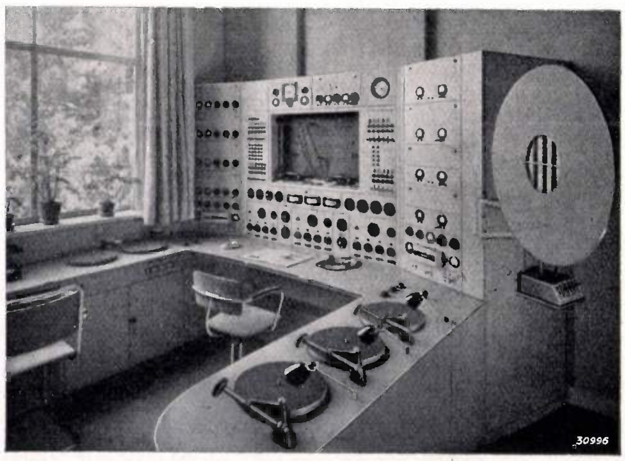


Fig. 1. A central control table in the old A.V.R.O. studio building. The person at this table had no visual contact with the studio. In addition to regulators and control instruments various auxiliary apparatus are installed on the table (gramophones, amplifiers, etc.).

[1]) The equipment of broadcasting studios, Philips techn. Rev. 4, 136, 1939.

[2]) See for example R. Vermeulen. The relationship between fortissimo and pianissimo, Philips techn. Rev. 2, 266, 1937.

elements and the necessary signalling systems, there is also a series of auxiliary apparatus, such as gramophone turntables, amplifiers, a switchboard for making the necessary connections between microphones, regulators, line to the transmitter, etc. This arrangement and the division of functions thereby involved between the collaborating technicians was chosen in order to permit as manysided use as possible of the control tables, for instance, for the broadcasting of radio plays in which different studios sometimes work together.

In the new studio building, which is intended only for the broadcasting of concerts, more emphasis could be laid on the collaboration between conductor and technicians. For each of the two studios (a large concert hall and a smaller studio for dance music) the mixing cabin and the control table are as it were combined, by bringing the control table itself into visual contact with the studio. At the same time the mixing and regulation and the whole direction of the broadcast are in the hands of a single person, who, by means of direct contact with conductor and orchestra, can follow the performance in all its details. The way in which the control tables for the two studios are arranged is shown in the floor plan of *fig. 2*, while *fig. 3* shows the control table which gives upon the large concert hall (*fig. 4*). In order to keep the view of the hall entirely free the table is low and only the operating elements are included, while the above-mentioned auxiliary apparatus of both tables are housed together in a separate room (the "technical service", see fig. 2). A more detailed description of the arrangement of the control tables is given in the text under fig. 3.

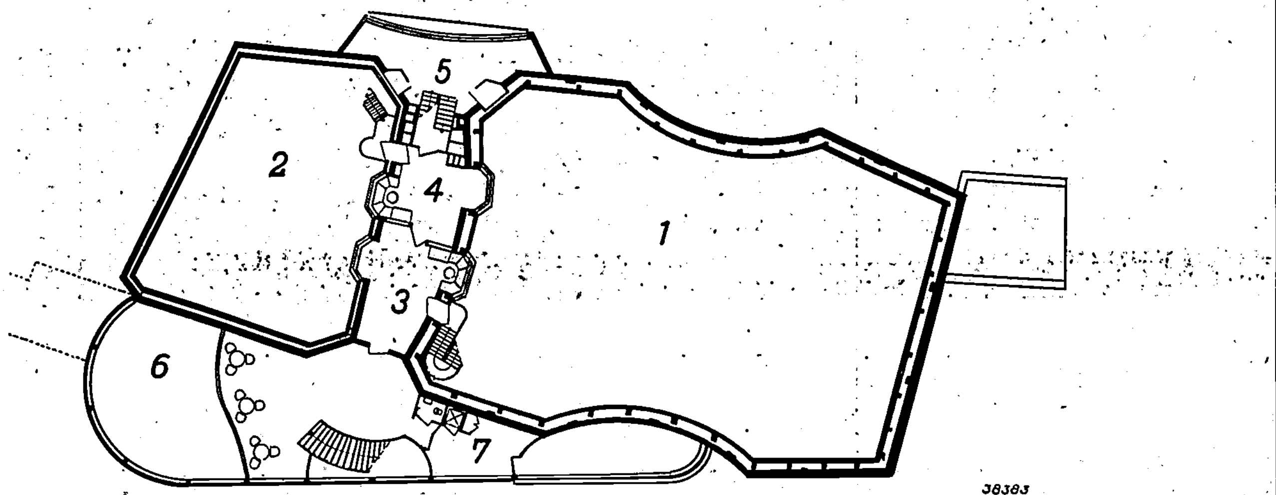


Fig. 2. Ground plan of the new A.V.R.O. studio building containing two studios: a large concert hall (*1*) and a smaller studio for dance music (*2*). Each studio has a control table. The two control tables are housed in cabins (*3, 4*) between the two studios and have a free view of the studios *via* triple glass windows which furnish adequate acoustic insulation. *5* technical service room, *6* hall, *7* conductor's room. The unusual floor plan of the large concert hall, which recalls somewhat the sound box of a violin, has been chosen for acoustic reasons. As may be seen the two studios are constructed as entirely free-standing boxes inside the building proper, with their own walls and a separate foundation. This separation between studio and outside world, which was also applied in the old A.V.R.O. building, and which has been carried through as far as possible, insures a very good acoustic insulation. (The figure is borrowed from "De 8 en Opbouw" **11**, 173, 1940).

The supply of the installation

Roughly speaking, the electro-acoustic installation of a studio building is nothing but a combination of a large number of amplifiers which must supply all kinds of low-frequency voltages at different levels to regulators, lines and loud speakers. For the general arrangement of a studio installation the article cited may be referred to [1]). Without entering into details we shall here consider a few essential points, the first of which is the supply.

The feeding of the many amplifiers, both with heating currents and anode voltages, usually took place from a central accumulator battery; this was the case in the old A.V.R.O. building. The battery there set up was, however, not large enough to supply the installation of the new building as well, so that the problem of supply had to be considered anew.

Now battery feeding of itself has several objecjections. Apart from the high cost of installation and the complicated maintenance, it is particularly the mutual coupling of the amplifiers connected

Fig. 3. Control table for the large concert hall. Due to the fact that all auxiliary arrangements not directly needed during the performance are housed elsewhere (in the Technical Service room, *5* in fig. 2), the table could be so arranged that an uninterrupted view of the hall is possible. On the five sloping panels are to be found from left to right: 1) switches and signal lamps for connections to the old building, a switch and a regulator for connecting the monitor loud speaker (upper right) to different studios, a microphone for giving directions to the studios, a set of knobs by means of which indications on a light screen can still be given to the conductor after the beginning of the broadcast. 2) A group of four microphone regulators for mixing four microphone contributions and a main regulator for regulating the whole. 3) Switches and signal lamps for signalling in the two studios and the corresponding announcer's cabins, and the indicating instrument of the modulation meter, with which the peak voltages in the programme are controlled. 4) Four more microphone regulators and a main regulator. 5) Telephone apparatus, switches for programme contributions from outside the studio, for instance for mixing sound effects (which can here be furnished from a separate studio of the old building), as well as keys and knobs for signalling to the second control table.

to the central battery which forms an objection. It is difficult to make the internal resistance of the battery so small that the coupling is sufficiently restricted, and recourse must therefore be taken to the introduction of decoupling circuits. Considering the very low frequencies (30 c/s) which must be dealt which in the amplifiers, the decoupling circuits must contain very large self-inductances [3]), which in turn involves considerable current surges upon switching on and off of an amplifier. These current surges lead to click disturbances in all the amplifiers connected.

In addition to these "natural" disadvantages of battery supply, in our case there was also the fact that it was a very uneconomical solution for the relatively limited installation of the new building to set up a second separate central battery. This was the reason why a different solution of the problem was applied, namely the supply of each amplifier separately by a built-in supply apparatus.

The fact that this solution, in which practically all coupling between the amplifiers is eliminated, has been applied until now only very seldom, is due mainly to the fear of the hum caused by the 50 c/s of the A.C. mains. This is indeed a difficulty which must not be underestimated, especially in a studio where all the amplifiers must be able to pass the frequency of 50 c/s in full strength, and where partially very weak signals occur which are therefore sensitive to disturbances. The problem could, however, be solved satisfactorily in the following way.

[3]) In telephone networks where the same problem occurs — cross talk between the microphone connected in parallel on a battery — the problem is much simpler, since practically only frequencies above 300 c/s are encountered.

Fig. 4. View of the large concert hall. Two hanging microphones on rails may be seen. Next to the position of the conductor is a movable signal box with a light screen on which the director of the broadcast, seated at the control table, can give several commonly occurring indications to the conductor. Right, sound effect loud speaker for the mixing of background noises. The numerous panels of the walls can be set at different angles in order to influence the acoustics of the hall (sound distribution, reverberation).

In order to limit the hum to the required degree the windings of input and output transformers of each amplifier must be especially protected against induction by spread lines of force from the supply transformer. Various means of doing this have been applied. In the first place the supply transformer works with a lower iron induction, *i.e.* not so close to saturation as is usually the case. This gives a higher permeability of the iron and therefore less spreading. Moreover, the supply transformer is placed in a heavy cast-iron pot which tends to absorb the spread lines of force and thus diminishes still more the field observable toward the outside. Furthermore the input and output transformer are placed in the amplifier panel as far as possible outside the sphere of the field of the supply transformer. This may clearly be seen in *fig.* 5: the supply transformer is at the upper left hand, the input transformer at the extreme right, the input transformer about in the middle at the bottom (because it is somewhat less sensitive to hum than the input transformer due to the greater signal intensity at that point). The input and output transformers are also housed in pots for shielding. Since it is here a question of shielding fields which are themselves quite weak, an alloy has been used for these pots which has a high permeability especially at low induction. Moreover, for the very sensitive input transformer an additional shielding inside the pot of a special kind of sheet metal is used, while its winding is "astatic", see *fig. 6*. By this means the sensitivity for external fields is made very small. An external field gives in general lines of force of the same direction in the two arms of the transformer core, in the two halves of the secondary winding therefore electromotive forces in the same direction are induced, and since the two halves are connected in opposition, the resulting voltage is zero. Nevertheless the transformer works quite normally for a primary current, because the latter causes a magnetic field whose lines of force run in opposite directions in the two arms of the core, so that the induced electromotive forces in the two halves of the secondary winding are added together.

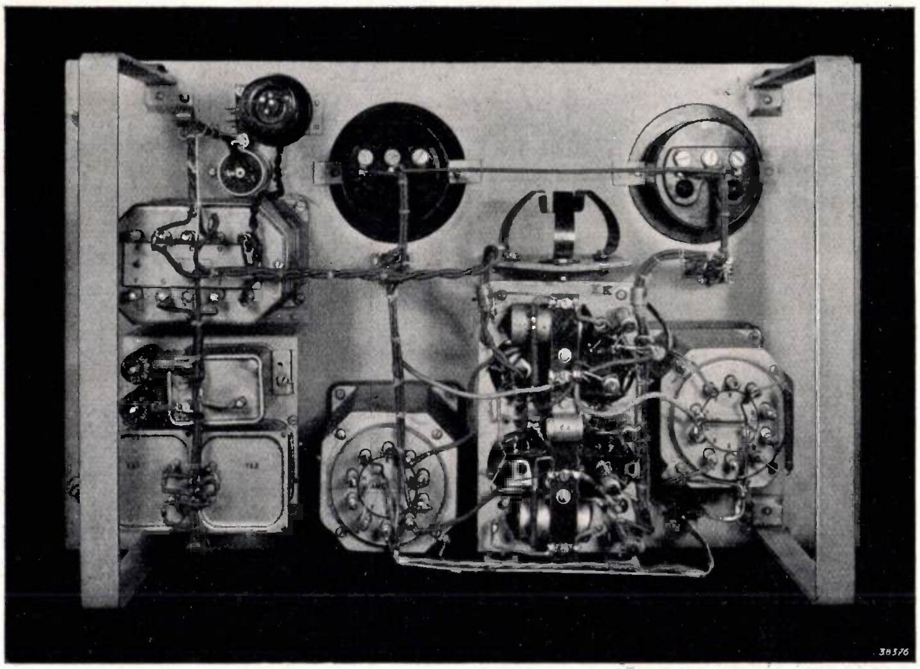


Fig. 5. Arrangement of the components in the panel of an amplifier. The supply transformer on the one hand (left) and the input and output transformer on the other hand (right and below) are kept as far as possible away from each other. Above the supply transformer may be seen a "Starto" valve (automatic starting resistance [4]), which limits the initial surge and thereby helps to prevent clicking in neighbouring amplifiers.

Special attention is also paid to the earth connection in order to decrease the hum. The transformer pots, which are to a certain degree coupled capacitatively with the windings, are not connected electrically to the panel, since this would enable the earth currents to follow all kinds of complex routes, but on the contrary they are insulated from the panel and connected to a common earth point of the amplifier by the straightest possible connections.

The cathodes of the amplifier valves are indirectly heated.

By these and other measures the hum could be sufficiently suppressed. A measurement of the hum and noise voltage [5]) in the case of the amplifier

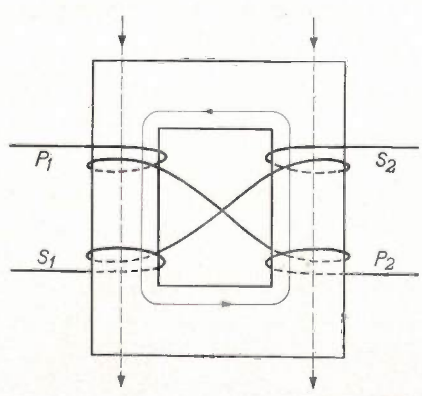


Fig. 6. Astatically wound transformer. P_1, P_2 halves of the primary, S_1, S_2 halves of the secondary winding. The primary current causes in the two arms of the core opposite lines of force, an external field, on the other hand, causes lines of force in the same direction (broken lines).

[4]) See P. C. van der Willigen, Philips techn. Rev. 1, 205, 1936.

[5]) The measurements were carried out with a so-called psophometer which takes into account the dependence on frequency of the sensitivity of the ear.

type *I* gave a value of 0.5 mV at the output, *i.e.* about 75 dB below the normal line level.

Standardization of the amplifiers

In older studio installations, including that of the old A.V.R.O. building, amplifiers of a different type were used for each function; thus there are microphone pre-amplifiers, gramophone amplifiers, programme amplifiers, coupling amplifiers, monitor amplifiers, etc. The differences between these types lie in the amplification produced, the input impedance, the output, etc.

This specialization and close adaptation of each amplifier to a certain purpose has disadvantages in manufacture as well as in use. In manufacture, because so many types of amplifiers must be made; in use, because so many amplifiers are needed, only a few of which are usually in action at one line, and for which the necessary reserves must be at hand. In our case, where each amplifier must also have a separate supply apparatus, these disadvantages were doubly felt.

Therefore a standardization of the amplifiers was undertaken for the new installation. All microphone connections, gramophone pick-ups, regulators are adapted to an input impedance of 200 ohms of the amplifier to be connected; by giving this an amplification sufficient for all purposes (60 dB), a single type of amplifier (*I*) proved sufficient. Only two other types of amplifier are then still needed for the listening to the broadcast at various spots with monitor loud speakers. The first is a coupling amplifier (*II*), which taps off the listening lines from the programme line (line to the transmitter) in such a way that commutations in the listening line have no reaction on the programme line, and which has a sufficiently high input impedance (5 000 ohms) not to affect the level of the programme line. The last is a power amplifier (*III*) which delivers the necessary power to supply a loud speaker (7 W). In fig. 10 at the end of this communication the way in which the three types of amplifiers are used may be seen.

All three types of amplifiers are two-stage push-pull amplifiers. By the push-pull connection all the connections can be kept symmetrical with respect to earth, which restricts the possible disturbances. Furthermore, due to the push-pull connection, the non-linear distortion is slight: the maximum distortion factor of the amplifiers of type *I* is 1 per cent above 100 c/s and 2 per cent below 100 c/s. In order to control the balance each amplifier stage contains a differential meter which indicates the difference current of the two valves.

Cable installation and cross connection panels

For the manufacture the desired simplification is simply obtained with the standardization of the amplifiers. In order to be able to take advantage of this in use also, amplifiers *I*, which can be used in various branches, are not included rigidly in certain lines, but the inputs and outputs of all these amplifiers are brought together on switchboards where all the desired connections can be made. The number of amplifiers of type *I* could thereby be limited to the relatively small number of 14, whereby each amplifier may act as a reserve for every other amplifier in case of defect. In *fig.* 7 may be seen the racks in which these 14 amplifiers are placed.

An important consequence of this centralization of the amplifiers is that the microphones in the studios now possess no pre-amplifiers situated in the immediate neighbourhood, but are connected by a relatively long connection to the amplifiers. This made it necessary to pay special attention to the combination of interferences with the weak microphone voltages. In the first place in designing the studio building more care had to be taken than in previous buildings that the routes of the

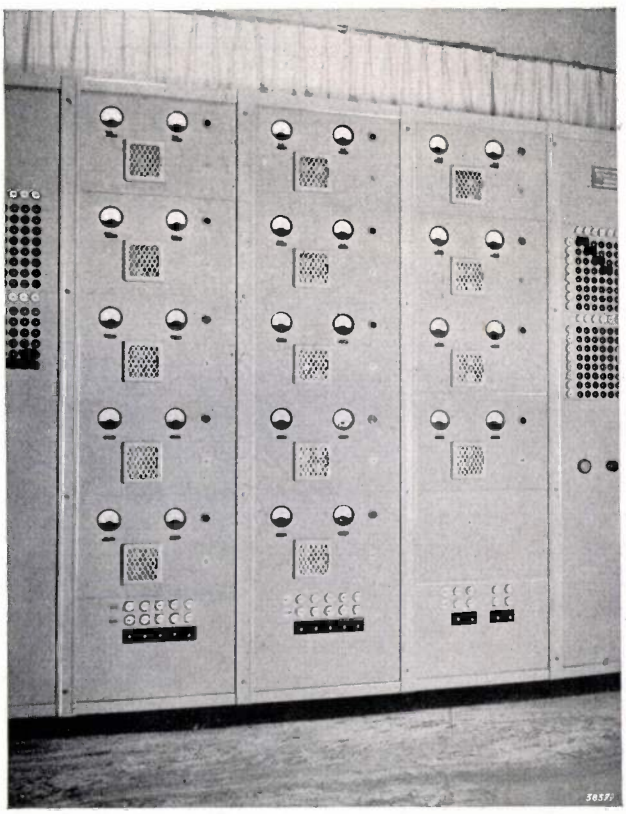
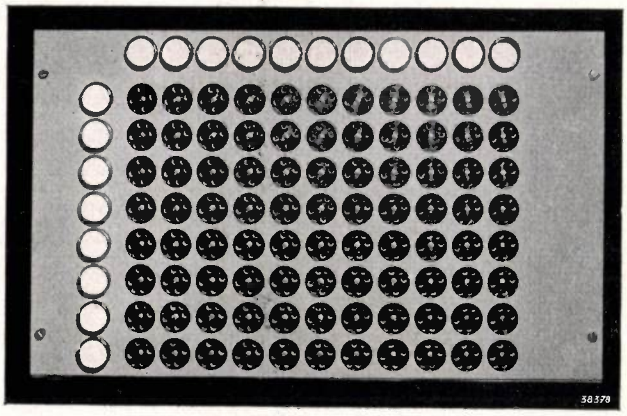


Fig. 7. Rack in the technical service room in which the 14 amplifiers *I* are placed. On the panel of each amplifier may be seen the two differential meters for controlling the balance, and a small door giving access to the valves.

a



b

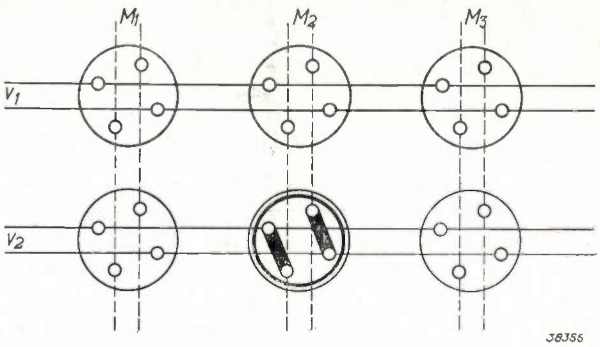


Fig. 8. *a*) Panel of cross connection units, with signal lamps at the ends of the horizontal and vertical rows.

b) Method of connecting the contact sockets of the cross connection units. M_1, M_2, M_3, etc., go to microphones and the like, V_1, V_2 etc. to amplifier inputs. A plug inserted for instance at the crossing point of V_2 and M_2 makes the connection there indicated.

"sensitive" microphone cables were kept away from those of the connections for the lights, the house telephones, the signalling, the electric clocks, etc. A plane at the height of the floor of the studios is therefore reserved exclusively for the microphone lines; the microphone connections are also at this height in the base boards of the studios. At places where the microphone lines must depart from this height, for instance for the connection with the points of suspension of the two hanging microphones in the large concert hall (fig. 4) a wide space is left free around the lines. All the connection lines are in continuous carefully earthed lead coverings for the sake of good shielding.

In the second place, the switching of the different microphone connections (14 in all) to the various amplifier inputs requires attention. The commonest method used in such cases is the switchboard, where each of the m outgoing connections is represented by a socket and each of the n incoming connections ends in a cord and a jack which can be plugged into any of the m sockets. This method could not be used here because of the very low level of the signals occurring at the amplifier inputs: the free hanging cords might here too easily lead to movements in the contacts and thus to crackle disturbances. A different system was therefore used, the cross connection panel. This can be pictured most simply as a set of m horizontal and n vertical rails crossing each other. Each horizontal rail is connected with an amplifier input, each vertical rail with a microphone line. By means of a jack a connection between two rails can be made at every crossing point.

Actually the two sets of rails are replaced by a panel of so-called "cross connection units", see *fig. 8*. Each connection unit contains two "horizontal" and two "vertical" contact sockets which are connected in the way shown in fig. 8*b*. The plugs which are inserted in these cross connection units (*fig. 9*) contain four very flexible contact pins of a special construction, which are connected in twos. In this way very good connections are obtained free of interference. In order to avoid mutual influences of microphones lines and amplifier inputs which are not connected with each other, there is a shielding in each cross connection unit between the "horizontal" and the "vertical" contact sockets. Furthermore the cross connection units also contain a number of auxiliary contacts (see fig. 9) which can be worked by two long auxiliary pins of the plugs and which may be used for signalling and locking. For instance, a series of signal lamps are operated by this means. These lamps are mounted around the cross connection panel (see fig. 8), when a plug is inserted at a crossing point the lamps at the ends of the horizontal and vertical rows involved are lighted to indicate the connection made and as a warning that these rows may no longer be used for other purposes.

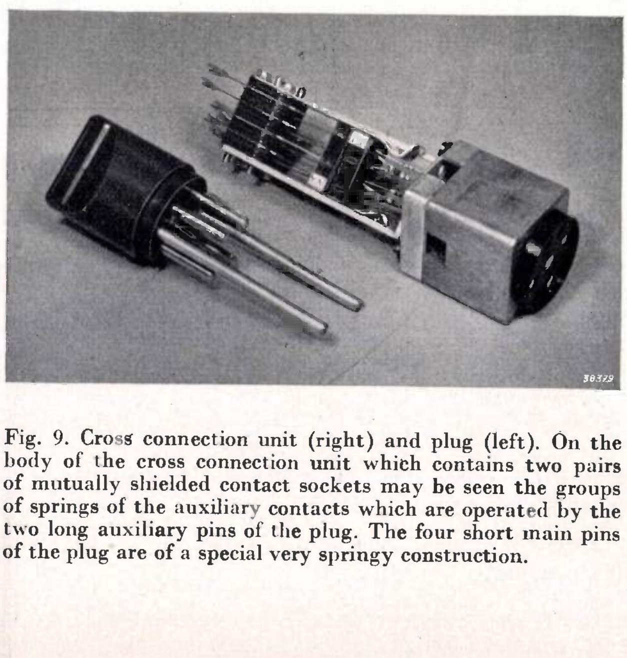


Fig. 9. Cross connection unit (right) and plug (left). On the body of the cross connection unit which contains two pairs of mutually shielded contact sockets may be seen the groups of springs of the auxiliary contacts which are operated by the two long auxiliary pins of the plug. The four short main pins of the plug are of a special very springy construction.

The outputs of the 14 amplifiers *I* also, with the lines to the regulators, the control instruments, the transmitter, etc. which must be connected to them are brought together on a cross connection panel. In fig. 7 it may be seen how the input and the output cross connection panels are placed to the left and right of the racks with the 14 amplifiers. The signalling system just described is so constructed that upon inserting a plug in the input cross connection panel the lamp at the corresponding horizontal row of the output cross connection panel also lights up faintly in order to facilitate the finding of the correct spot for plugging in on that panel, and at the same time to indicate again that the amplifier in question is in use.

In conclusion in *fig. 10* a view of the complete installation is given. On the cross connection panels as an example the connections are made for a broadcast in which three microphones in the studio collaborate (main regulator H_1), while an acoustic background (sound effect) is provided by the gramophone turntable which can be added to the programme *via* the main regulator H_2. At the same time this background can be sent to the studio itself *via* the regulator *R*. The communications of the announcer can also pass over the main regulator H_2. Several details are further explained in the text under fig. 10.

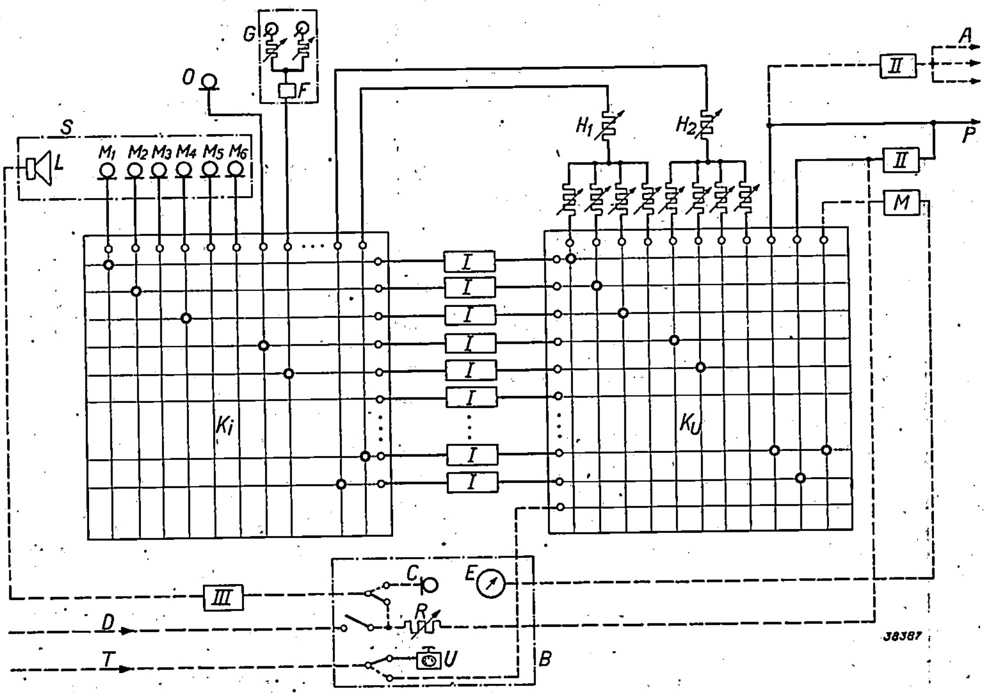


Fig. 10. View of the electro-acoustic installation of the new studio building. K_i input cross connection panel, K_u output cross connection panel, *S* studio with six microphones M_1-M_6 and sound effect loud speaker (at the same time loud speaker for giving directions); *O* microphone in the announcer's cabin, *G* gramophone turntable with correction filter *F*, H_1 and H_2 main regulators, *P* programme line, *A* monitor lines, *I* universal amplifier, *II* coupling amplifier, *III* power amplifier, *M* modulation meter with indicating instrument *R*; *B* cabin with microphone for giving directions *C*, regulator *R* for sound effect which can be supplied over line *D* from the old building, *U* telephone apparatus connected to the telephone line *T*, over which contributions to the programme may also be brought in.

THE CALCULATION OF LIGHTING INSTALLATIONS WITH LINEAR SOURCES OF LIGHT

by H. ZIJL.

535.241 : 628.93

Several diagrams are derived which make it possible to determine in a simple way the illumination, and from it the luminous efficiency for installations equipped with linear light sources.

Introduction

In designing a lighting system with ordinary electric lamps and normal fittings one can make use of extensive factual material, of a theoretical as well as of an empirical nature, about the distribution of the illumination with the customary arrangement of the light sources. In the literature more or less detailed information can be found, which is usuallly arranged in the form of tables[1]) and in general is sufficiently accurate for practical use. The modern linear sources of light, however, such as the "Philora" TL lamps. are entirely different from the lamps generally used until now with respect not only to the distribution of their luminous flux but also to the way in which they are installed in the rooms to be illuminated, so that the experience which has been gained with earlier lighting systems cannot be directly applied to them. The number of systems already installed with such linear light sources is still too small to make it possible to collect adequate experimental data from them. It is therefore necessary for the present to calculate the distribution of the luminous flux for linear light sources theoretically, as was also formerly the case for ordinary systems with ordinary electric lamps. From this the so-called luminous efficiency of the system can easily be found. This is of course a more elaborate method than consulting a table, but it is unavoidable until such practical tables have been crystallized out of many calculations confirmed by illumination measurements. Moreover, it offers the opportunity of attaining a clearer insight into the problems in question than is usually obtained by the mechanical calculation with tables.

Illumination systems with linear light sources usually consist of tubes which are mounted horizontally close under the ceiling with or without reflectors. Such a system must then provide the greatest possible illumination on the tables present in the room, which generally have their working surface about 1 m above the floor. The luminous flux from the lamps may be divided into three parts (*fig. 1*) which are incident:

a) directly on the working surface,

b) directly on the ceiling or the inner surface of the reflector and

c) on the walls of the room:

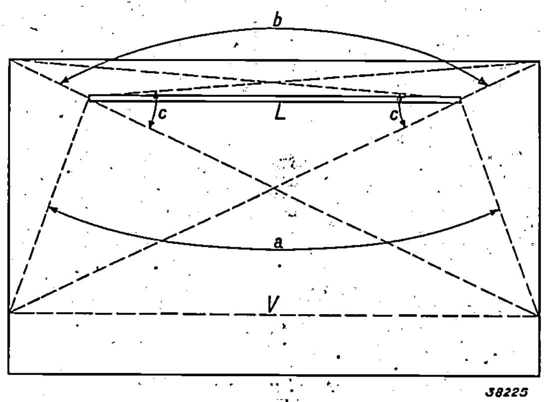


Fig. 1. The linear light source L is placed against the ceiling. The radiation a strikes the working surface V directly, while b is directed toward the ceiling and c toward the walls.

On the working surface, in addition to the direct radiation a), is also incident

1) part of the flux which has been reflected once by the ceiling or by the reflectors in which the lamps are placed,

2) part of the flux which has been reflected once by the walls of the room and

3) part of the flux which has been reflected repeatedly between the walls and the ceiling or between the various walls.

It is in general therefore permissible in such considerations to assume that only the part of the room which is higher than the working plane takes part in the repeated reflections, while the contribution from surfaces which are in or below the working surface may be neglected.

Of the four different contributions to the total illumination of the working plane which have been mentioned in the foregoing, the direct illumination will in general be by far the most important. This contribution should therefore be calculated accurately within reasonable limits. The other

[1]) Such tables are as a rule based upon the original investigations of W. Harrison and E. A. Anderson, Trans. I.E.S., 11, 67, 1916.

contributions are usually successively smaller in the order in which they are mentioned, and may therefore be calculated less accurately without causing important errors in the final result. Since, moreover, the calculation of these smaller contributions is fundamentally the same as in the case of ordinary lighting systems, or otherwise the same as that which will be given in this article for the direct illumination produced by the linear light sources, we shall here concern ourselves chiefly with the calculation of the direct illumination intensity.

The calculation of the direct illumination with linear sources of light

In this calculation of illumination provided by linear light sources we shall begin with the simple considerations which have already been given in this periodical [2]) about the light distribution of these sources. We assume that the surface of the TL lamp (*fig. 2*) radiates diffusely according to

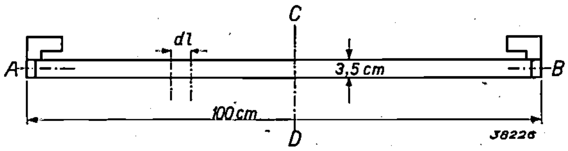


Fig. 2. Linear light source TL 100 with a length of 100 cm and a cross section of 3.5 cm. AB is the axis of the lamp and CD a perpendicular plane through the middle of the lamp, while a linear element of the lamp is indicated by dl.

Lambert's law. Observed from a sufficiently great distance the linear lamp then behaves as a point source would do which had a luminous flux distribution of the form indicated in *fig. 3*. Perpendicular to the axis of the lamp the maximum luminous flux of I_m c.p. is radiated, and in the direction of the axis of the lamp the light flux is zero, while in directions which make an angle of 90-α with the axis of the lamp the flux is given by:

$$I_\alpha = I_m \cos\alpha \quad \ldots \ldots \quad (1)$$

As may be seen in fig. 3, the distribution of the luminous flux is now represented by a circle with a diameter I_m which is tangent to the axis of the lamp AB and revolved around it; *i.e.* the light intensity is a torus with no hole.

The light flux F (lumens) which is emitted by a point source of light with an isotropic (spherical) luminous flux distribution of I c.p. is given by:

$$F = 4\pi I.$$

[2]) N. A. Halbertsma and G. P. Ittmann, Philips techn. Rev., 4, 181, 1939.

The total luminous flux F which is emitted from a light source with a torus-shaped light intensity distribution according to fig. 3, becomes, however,

$$F = \pi^2 I_m.$$

For a linear light source with a length l one

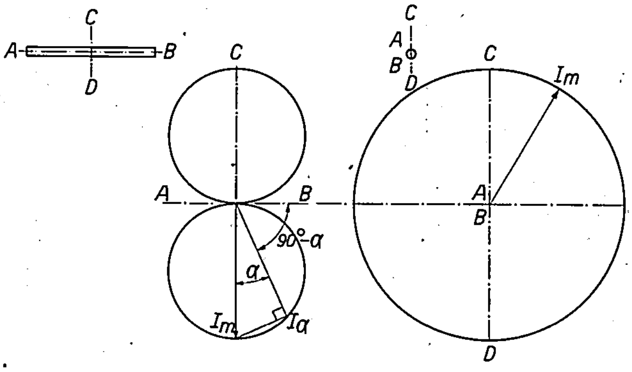


Fig. 3. Distribution of illumination from the linear light source AB in a plane through the axis of the lamp and in the perpendicular bisecting plane, respectively. I_m is the maximum luminous flux and I_α is the luminous flux at an angle 90°—α with the axis of the lamp AB. The spatial distribution of the luminous flux is obtained by revolving the lefthand figure around AB; it is thus a torus with no hole.

may speak of the luminous flux F_0 which is radiated per unit length, so that:

$$F = F_0 l$$

The linear lamp may thus be considered as a row of elements of unit length each having a maximum luminous intensity of

$$I_0 = \frac{I_m}{l} \quad \ldots \ldots \quad (2)$$

The luminous flux per unit length is then

$$F_0 = \pi^2 I_0 \quad \ldots \ldots \quad (3)$$

while the total luminous flux radiated by a linear light source is given by

$$F = \pi^2 I_0 l.$$

On the basis of equations (1) and (2) it is now immediately clear that the contribution dI_α furnished by a cylindrical element of surface dl at the point P of a linear light source (*fig. 4*) to the luminous intensity in the direction PO making an angle of 90°-α with the axis of the lamp will be:

$$dI_\alpha = I_0 dl \cos\alpha \quad \ldots \ldots \quad (4)$$

The contribution dE_α of this to the illumination (expressed in 10^4 lux, since we calculate with cm instead of m) at point O on the plane $ABCD$ situated at a distance h cm from the linear light source is given by:

$$\mathrm{d}E_\alpha = \frac{\mathrm{d}I_\alpha}{r^2}\cos\gamma = \frac{I_0}{r^2}\,\mathrm{d}l\,\cos\alpha\,\cos\gamma,$$

where r represents the distance OP in cm, and γ the angle between the direction of radiation PO and the vertical to the irradiated plane $ABCD$. If we call the perpendicular distance from the point O to the linear light source a (cm), then the contribution to the illumination at O on the plane $ABCD$, furnished by a cylindrical element of the linear light source which is observed from O at an angle $\mathrm{d}\alpha$, may be written as follows:

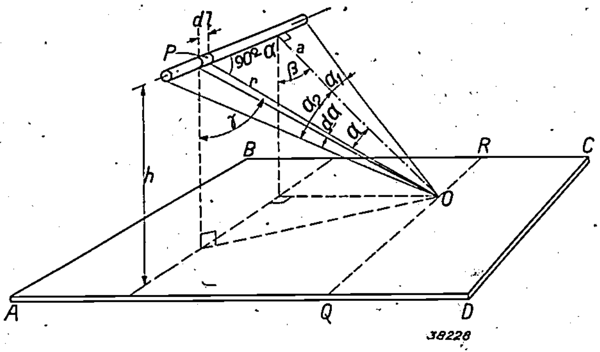


Fig. 4. Sketch showing the contribution to the illumination at the point O of the line QR on the working surface $ABCD$ which is furnished by the linear element $\mathrm{d}l$ at point P of the linear source.

$$\mathrm{d}E_\alpha = I_0\,\frac{h}{a^2}\cos^2\alpha\,\mathrm{d}\alpha \quad \ldots\ldots \quad (5)$$

The total illumination E which the linear light source produces at O on the plane $ABCD$ may now be found by integrating this expression (5) between the limits α_1 and α_2 keeping in mind that in fig. 4 the angle α_2 is positive and we must therefore take α_1 negative, because for the perpendicular a from O on the linear light source the angle $\alpha = 0$. We then obtain the formula:

$$E = I_0\,\frac{h}{a^2}\cdot\frac{1}{4}\left|2\alpha + \sin 2\alpha\right|_{\alpha_1}^{\alpha_2}\cdot 10^4\ \text{lux}, \quad (6)$$

where α must be expressed in radians, and h in cm, and I_0 represents the luminous intensity in c.p. per cm.

We shall consider this general formula (6) somewhat more closely for a simple case in which it will be found that it actually does give the familiar formula. If we consider the illumination E at distances a which are small compared with the total length l of the linear source, then

$$\alpha_1 = -\frac{\pi}{2} \text{ et } \alpha_2 = +\frac{\pi}{2} \text{ so that we obtain}$$

$$E = I_0\,\frac{h}{a^2}\,\frac{\pi}{2}\cdot 10^4\ \text{lux} = \frac{\pi}{2}\,\frac{I_0}{a}\cos\beta\cdot 10^4\ \text{lux},$$

where β represents the angle γ for $\alpha = 0$. For a surface element which is perpendicular to the shortest line joining it with the linear source, $\beta = 0$, so that we then obtain simply that the illumination varies inversely proportionally with the distance:

$$E = \frac{\pi}{2}\cdot\frac{I_0}{a}\cdot 10^4\ \text{lux} = 1{,}57\,\frac{I_0}{a}\cdot 10^4\ \text{lux},$$

which is even more easily understood on the basis of simple considerations of symmetry, as given earlier in this periodical [3]).

Diagrams for the determination of the direct illumination

In practical cases in order to be able to find the direct illumination with linear light sources easily and rapidly, the most obvious method is to attempt to reduce the general formula (6) to a simple diagram.

In *fig. 5* the plane is drawn through the linear source and the point O at which the illumination on the working surface is desired. It may immediately be seen that the total illumination E produced at O by the source of length l can be made up of the contribution E_2 of the light source l_2 and

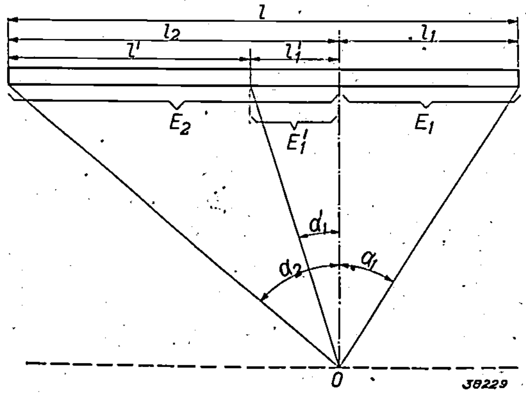


Fig. 5. The plane through the linear light source l and point O, at which the illumination on the working surface is to be determined, is here drawn to scale. The parts l_1, l_2 and l_1' of the linear source, which are seen from O at the angles α_1, α_2 and α_1', respectively, furnish the contributions E_1, E_2 and E_1', respectively, to the total illumination at point O on the working plane.

the contribution E_1 of l_1, while the illumination given by a light source l', on whose extension the base of the perpendicular from O lies, is of course the difference between the illuminations E_2 and E_1', which would be produced at O by the light sources l_2 and l_1' respectively.

This is all expressed (with correct sign) in formula (6) when we take α_2 and α_1' positive and α_1 negative, because the expression $^1/_4\,(2\alpha + \sin 2\alpha)$ becomes zero for $\alpha = 0$, so that, for example, the illumination E_2 given by the lamp l_2 at O is given by:

---

[3]) Philips techn. Rev. 4, 181, 1939.

$$E_2 = I_0 \cdot \frac{h}{a^2} \frac{2\alpha_2 + \sin 2\alpha_2}{4} \cdot 10^4 \text{ lux.} \quad . \quad (7)$$

This illumination E_2 at a point O of the line QR on the working surface $ABCD$ in fig. 4 can now be plotted in polar coordinates from O at an angle α_2 with the perpendicular to QR, as has been done in principle in *fig. 6*. In this diagram we have plotted

$$e = \frac{2\alpha + \sin 2\alpha}{4} \cdot 10^4. \quad . \quad . \quad . \quad . \quad (8)$$

In order to obtain the required illumination E_2, therefore, it is only necessary to multiply the value of e_2 at an angle of α_2 read off in fig. 6 by I_0/ha^2[4]).

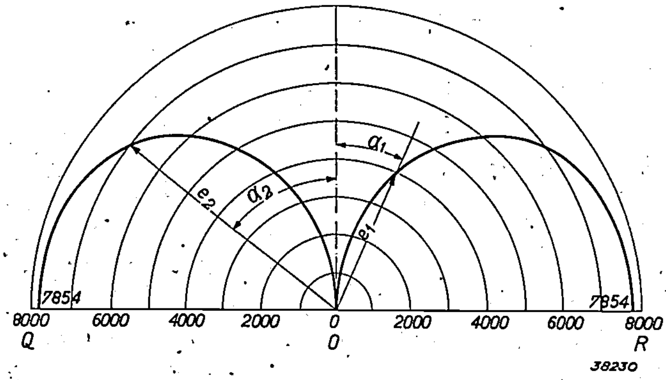


Fig. 6. The illumination E which a part of a linear light source with a luminous flux of I_3 c.p./cm produces at a point O which makes an angle α with the perpendicular distance a to the lamp amounts to I_3h/a^2 times the quantity e plotted in this polar diagram, when h is the distance from lamp to working surface in cm.

For the "Philora" lamps TL 100 which have a luminous flux F_0 of 10 lumens per cm, according to formula (3) we may assume about 1 for the candle power I_0 per cm. For "Philinea" lamps I_0 amounts to about 0.6.

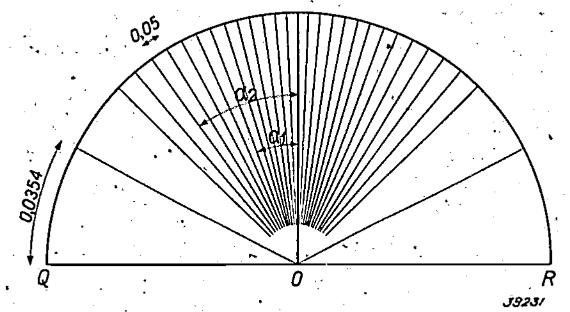


Fig. 7. In this polar diagram sectors are plotted each of which furnishes a contribution of 0.05 to the expression

$$\frac{1}{4} \left| 2\alpha + \sin 2\alpha \right|_{\alpha_2}^{\alpha_1}$$

A part of the lamp which, as seen from O, extends over n sectors furnishes a contribution to the illumination at O of 500 n I_0h/a^2 lux.

4) In illuminating engineering the distances a and h are usually expressed in metres. The value of e in this polar diagram must then of course be decreased by a factor 100.

Usually the angles α_1 and α_2 will be so large that the corresponding values of e_1 and e_2 are easily read off if fig. 6 is simply laid over fig. 5 which is drawn to scale. For rays which pass through O at a small angle α to the perpendicular to QR, however, their point of intersection with the illumination diagram in fig. 6 is difficult to determine sufficiently accurately, and for these cases it is therefore advisable to use the diagram of *fig. 7*. This diagram is obtained by plotting sectors polarly from O each of which contributes the same amount to the direct illumination intensity on the working surface in the neigbourhood of O. For each sector of fig. 7.

$$\left| \frac{2\alpha + \sin 2\alpha}{4} \right|_{\alpha_1}^{\alpha_2} = 0{,}05 \quad . \quad . \quad . \quad . \quad (9)$$

so that between $\alpha = 0$ and $\pi/2$, for which expression (9) becomes 0 and $\pi/4 = 0.7854$, respectively, lie fifteen complete sectors, while for the sixteenth and last sector expression (9) amounts to only 0.0354. If fig. 7 is sufficiently carefully drawn, the illumination at O can then be derived from it with a certain degree of accuracy, especially for fairly small angles α, in which region the successive sectors are almost equal in size, by laying fig. 5 drawn to scale over it. If, for example, the number of sectors in the diagram of fig. 7 compassed by the linear light source of fig. 5 seen from O is n, then according to formula (6) the direct illumination which it produces at O will be represented by:

$$E = 500 \; n \; I_0 \frac{h}{a^2} \text{ lux} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Practical calculation of the distribution of the illumination intensity over the working surface

On the basis of a simple example we shall now explain somewhat more precisely how with the help of figs. 6 and 7 the direct illumination intensity can be determined at a sufficient number of points on the working surface to furnish data for a rough estimation of the illumination efficiency of a system of linear sources of light. We consider a room of 13 by 12 m with a height of 4 m. In this room 64 TL 100 lamps 1 m in length are placed against the ceiling in 8 parallel rows of 8 lamps each. These lamps are about 3 m above the working plane and at distances of 1.5 m from each other. On the working surface (*fig. 8*) we now draw the straight lines *II*, etc. up to *IX* vertically beneath the rows of lamps. It is along these lines that we wish to know the illumination intensity. At 1.5 m outside the lines *II* and *IX*, respectively, there are also drawn the lines *I* and *X*, both of which thus lie 25 cm outside the room. By allowing the distances between the lines to depend in a simple way upon those between the rows of lamps, provision may be

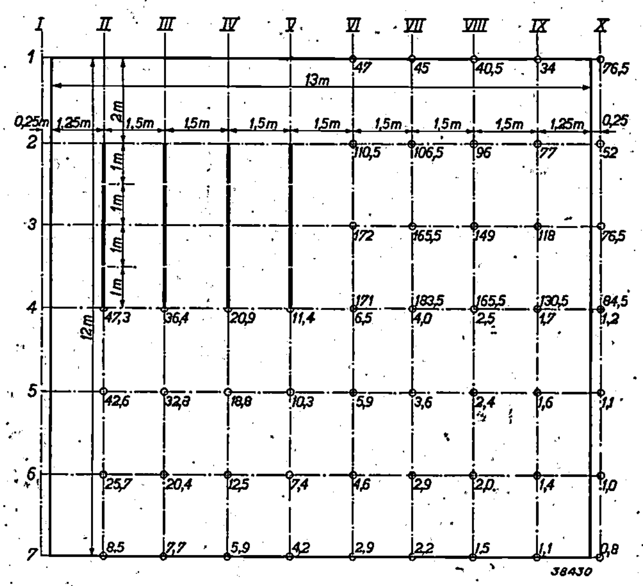


Fig. 8. In a room of 12 by 13 m and 4 m high there is a total of 8 times 8 linear lamps type TL 100 installed as indicated for the upper left-hand quadrant. In the lower half of the figure the contribution to the illumination of the row of lamps *II* is given for the intersection points of the working lines. In the upper right-hand quadrant the total illumination intensity is given for each of the points of intersection.

made that the successive lamps always give the same contribution to the illumination intensity at corresponding points on the successive lines. By a suitable choice of the working lines we thus simplify the calculation considerably.

We also draw several working lines in a direction perpendicular to the rows of lamps, and we have drawn them here 2 m apart so that there are seven of them (numbered from *1* to *7*) including the walls of the room. With the help of the diagrams of figs. 6 and 7, we now calculate for the points of intersection of these two systems of lines the contribution to the direct illumination furnished by every lamp. In the upper left-hand corner of fig. 8 the position of the lamps is indicated and in the lower half the contribution for every intersection is given which the row of lamps *II* furnishes to the total direct illumination of the working surface at that point. In the upper righthand corner of fig. 8 the direct illumination is noted for each of the intersections, rounded off to $^1/_2$, as furnished by the whole system at that point. The variation of this direct illumination in the other three quadrants is of course a mirror image of that in the quadrant calculated.

Along two mutually perpendicular lines *2* and *IX* in the upper right-hand quadrant we have now in *fig. 9* indicated the variation of the direct illumination. From this variation the points can immediately be found at which the direct illumination amounts to a whole number multiple of 10 lux. The mirror images of these points in the quadrant of the working surface lying to the left above in fig. 9 can thus be indicated without difficulty, and the lines may be drawn through them indicating equal direct illuminations, which give us the so-called isolux diagram for the upper left-hand quadrant of the working surface. Although the direct illumination in the middle of the room does not vary rapidly, even when the lamps are regularly spaced it falls off rapidly near the walls, and in the direction of the axes of the lamps it even falls to less than $^1/_3$ of its value in the middle of the room, as will be clear from fig. 9. If the greatest possible uniformity in light distribution over the whole working surface is desired, therefore, the light sources should be more closely spaced along the edges of the ceiling than has been done in this example.

If the luminous efficiency of a system is desired the total light flux which strikes the working surface should be determined, because the luminous efficiency is equal to this quantity divided by the luminous flux installed. From the illumination figures of fig. 8 by multiplication by the corresponding surface elements the total direct luminous flux which strikes the working surface can be relatively easily determined. For the quadrant under consideration with a surface of 39 cm² the direct luminous flux is 4 100 lumens, so that the average direct illumination will be about 105 lux. The ceiling and walls of the room to be illuminated have a reflectivity of 35 per cent. On the basis of the tables mentioned in the introduction (*cf.* footnote [1])), it can now be calculated that about 20 per cent of the light flux directed toward the ceiling finally reaches the working surface, while of the luminous flux directed toward the walls about 12 per cent also becomes effective. The total installed luminous flux is 64 000 lumens; of this nearly half is directed toward the ceiling, namely about 31 000 lumens. Of this half therefore 20 per cent finally reaches the working surface, *i.e.* about 6 200 lumens. Since the direct luminous flux to the whole working surface of 156 m² is $4 \times 4\,100 = 16\,400$ lumens and 31 000 lumens go to the ceiling, 16 600 lumens are directed toward the walls. Of this, 12 per cent finally reaches the working surface, *i.e.* 2 000 lumens. Due to the reflections at ceiling and walls, therefore, a luminous flux of $6\,200 + 2\,000 = 8\,200$ lumens reaches the working surface of 156 m² and provides on it an increase of the average illumination of about 55 lux. This makes the average of the total illumination over the working surface $105 + 55 = 160$ lux. It must, however, be taken into account that after the passage of time the reflectivity of ceiling and walls will decrease, as well as the light flux given by the lamps. It is therefore advisable to count only on a value of 140 lux for the average illumination intensity. The total "effective" luminous flux which then strikes the working surface of 156 m² will now be about 22 000 lumens. Since the total light flux installed is 64 000 lumens, this means a luminous efficiency of nearly 35 per cent.

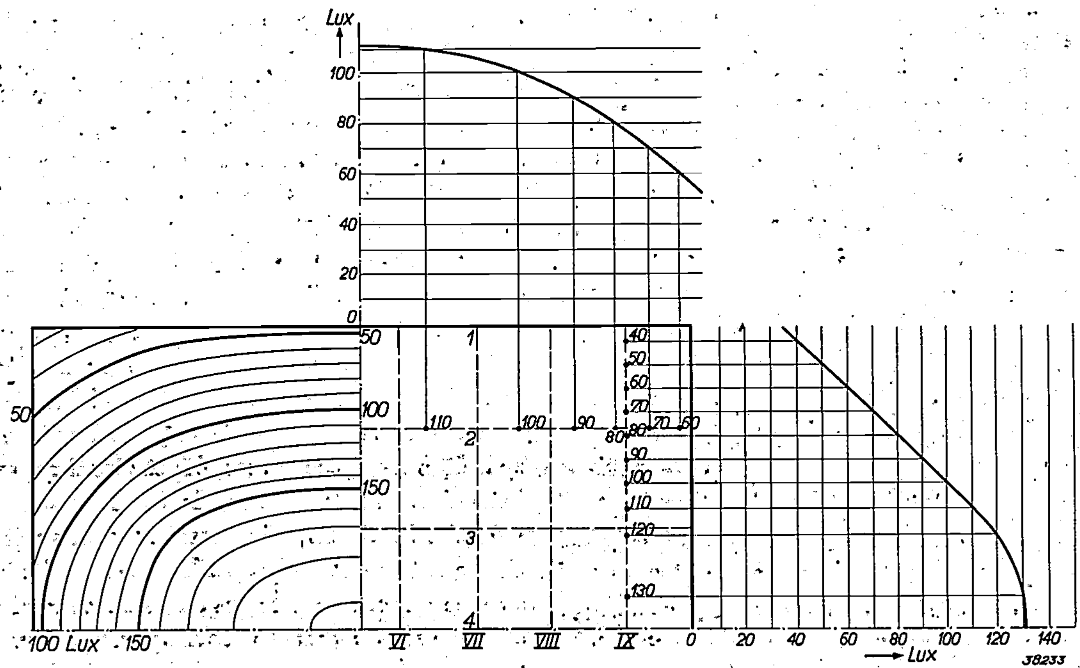
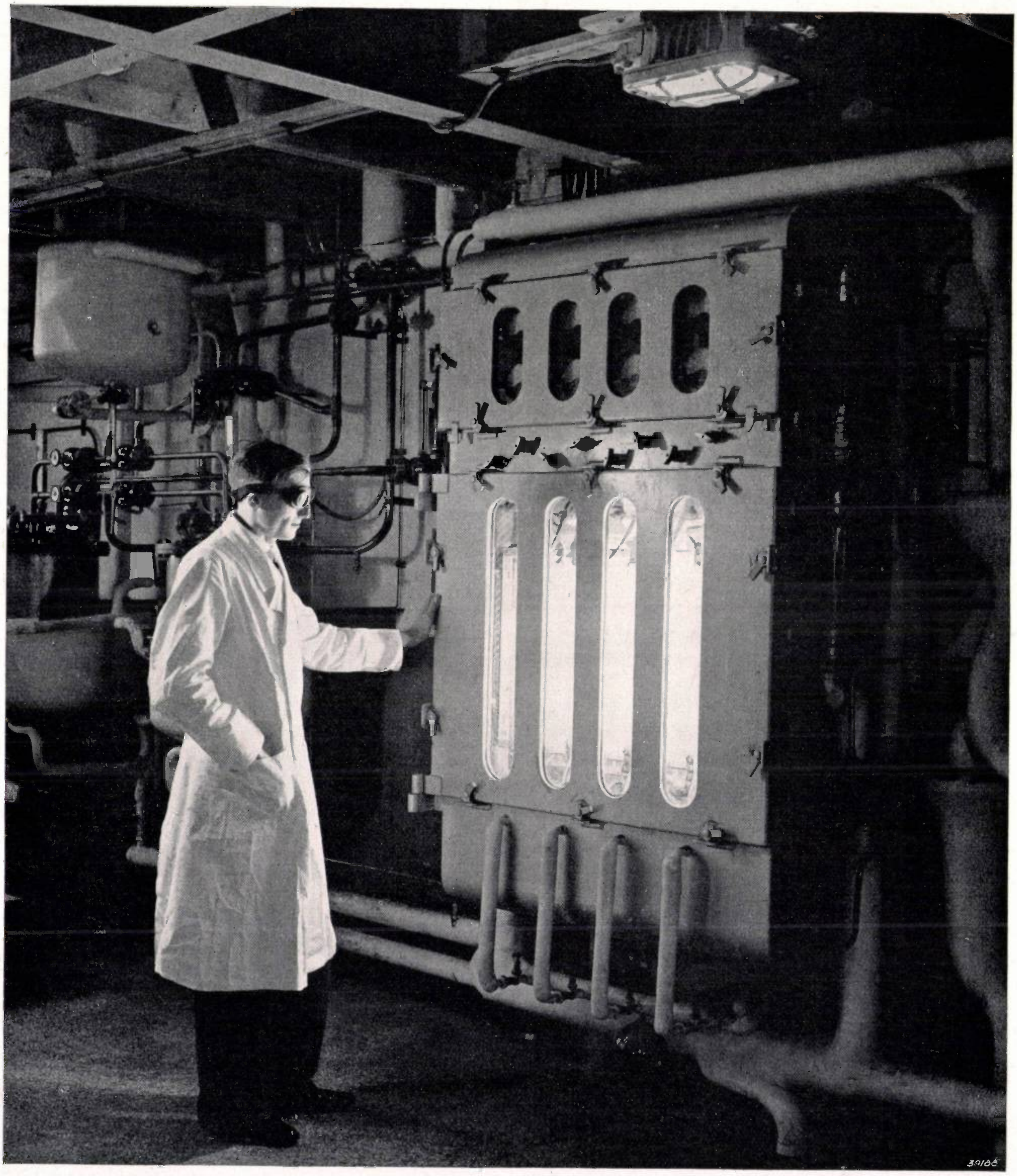


Fig. 9. For the working lines 2 and *IX* in the upper right-hand quadrant of fig. 8 the variation of the illumination is here drawn. The points along these working lines in which the illumination amounts to a whole number multiple of 10 lux are found graphically, and their mirror images can thus be drawn in the upper left-hand quadrant, for which we can in this way determine the isolux diagram of the direct illumination. In order to obtain a more uniform distribution of the illumination over the working surface it is found desirable to concentrate the lamps somewhat more toward the edges of the ceiling than was done in this example.

PREPARATION OF VITAMINE D



In the works of N.V. Philips-van Houten at Weesp the vitamine D is obtained from provitamine in this irradiating apparatus, which contains four high-pressure mercury lamps each consuming 1·5 kW.

A SIMPLE HIGH-FREQUENCY OSCILLATOR FOR THE TESTING OF RADIO RECEIVING SETS

by J. D. VEEGENS and M.K. DE VRIES.

621.396.615.1:621.317.7

For the testing and adjustment of radio receiving sets after manufacture, as well as for purposes of service, a measuring oscillator has been designed which gives a high-frequency voltage whose amplitude is adjustable continuously between 1 μV and 0.1 V, while the frequency can be varied continuously between 100 000 c/s and 60 Mc/s (wave length 5 m). By means of a built-in low-frequency oscillator the amplitude of the oscillator signal can be modulated by 400 c/s. In this article the fundamental problems are discussed which are encountered in the construction of the measuring oscillator, particularly that of keeping the frequency stable, regulating the amplitude and preventing frequency modulation upon modulation of the amplitude of the oscillator signal. In conclusion several details of the construction of the apparatus are discussed.

For the testing and adjustment of radio sets an oscillator is needed which can deliver to the receiving set voltages, which have the same frequency and amplitude as the aerial voltages when the set is in use. In some cases a sinusoidal signal of constant amplitude will be enough, while in other cases the signal should be modulated in amplitude like the aerial voltages.

Upon closer examination of the possibilities of application of such an oscillator, it is found that the requirements which must be made of the apparatus may vary very much according to the nature of the application. For special measurements in laboratories and factories it may be necessary to know the frequency or the amplitude of the oscillator signal very accurately, or the form of the signal must be exactly that of a sine. In by far the majority of applications, however, less severe requirements suffice. In practice, precision measurements are not usually carried out with the oscillator, but only more or less qualitative tests. This is particularly the case for all tests which are carried out in service stations for detecting defects in radio sets. For such tests an oscillator signal of reasonably constant amplitude and frequency is also needed. The value of the frequency need not, however, be able to be adjusted with greater accuracy than about 1 per cent, while as to the amplitude, it is only necessary to know its order of magnitude.

In the following we shall describe a simple measuring oscillator (type GM 2 882) which is suitable for all applications where the highest requirements are not made. In order to lend the apparatus universal utility special care was taken that the frequency and the amplitude of the oscillator signal could be varied within very wide limits. Furthermore, emphasis was laid on practical features, such as easy operation and a strong and light construction of the apparatus. The measuring oscillator can easily be carried, which is of special importance for service purposes in order to be able to study faults in receiving installations on the spot.

Several fundamental problems

The problems which were encountered in the development of this apparatus may be divided into three groups:

1) Securing an oscillator frequency which can be varied within wide limits and which remains absolutely constant upon each adjustment of the measuring oscillator.
2) Securing an output voltage whose amplitude can be adjusted in a reproducible manner in a region with a lower limit of about 1 μV.
3) Securing the desired amplitude modulation of the signal.

The solutions of these three problems attained in our apparatus will now be dealt with successively.

Adjusting the frequency and keeping it constant

The frequency region of the measuring oscillator extends from 100 000 to 60 Mc/s, corresponding to wave lengths of 3 000 to 5 m. The lowest frequencies are intended for the investigation of the intermediate-frequency part of receiving sets, while the highest frequencies may be used for the investigation of the ultra short wave part of radio receivers. The adjustment of the frequency takes place in the usual manner by continuous regulation of a condenser and commutation of a self-induction coil. Six frequency regions may be chosen in this way, which are distributed as follows:

10— 30 kc/s
30—100 kc/s
100—300 kc/s
3— 10 mc/s
10— 30 mc/s
30— 60 mc/s

The principles of the connections of the oscillator are given in *fig. 1*. The frequency of the oscillations

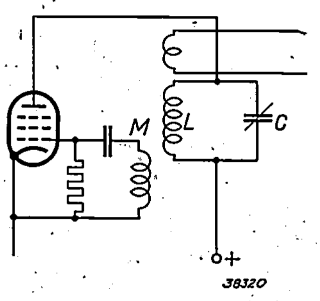


Fig. 1. Principle of the connection of the measuring oscillator. The oscillation circuit is not connected in the grid circuit, but in the anode circuit, since in this way a more nearly constant oscillator frequency is obtained.

excited is determined by the oscillation circuit LC according to the familiar equation:

$$f = \frac{1}{2\pi\sqrt{LC}} \quad \ldots \ldots \quad (1)$$

Equation (1) is actually only an approximation. There are numerous causes which result in a small displacement of the frequency, since the exact value of the frequency depends upon the resistances, capacities and self-inductions of all components which are coupled with the oscillation circuit, as well as upon the slope of the oscillator. These frequency shifts would of themselves be no objection if they could be taken into account upon calibrating the apparatus. Several of the quantities which affect the oscillator frequency, especially the loading resistance and certain capacities, exhibit, however, the undesired property that they may change with the conditions of performance. In order to prevent the occurrence of unallowable fluctuations in the oscillator frequency from these causes, very great attention had to be paid to these phenomena in the construction of the apparatus.

The influence of variations in the external loading was practically completely eliminated by not connecting the oscillator stage directly to the apparatus to be tested, but including an amplifier stage behind the oscillator stage. Variations in the working state of the amplifier stage will not react appreciably on the oscillator stage, since the two are only very loosely coupled. This will be dealt with later on.

We shall now go somewhat more deeply into the causes of capacity variations. The chief working conditions which may effect the capacities are the temperature of the surroundings and the supply voltage. The temperature of the surroundings affects mainly the dielectric constants of the insulation materials, whereby the capacities also experience a certain change. The thermal expansion of the tuning elements also results in a certain change in capacity, and, moreover, a change in the self-induction. It was, however, possible to make the temperature coefficients of the tuning elements so small that the temperature variations occurring in practice cause no unallowable fluctuations of the frequency. By suitable placing of the components, moreover, provision was made that the internal heating of the LC circuit during use is not too great. The result of these measures is that the frequency changes about 1 ‰ at the most due to the heating of the apparatus during use.

The influence of the s u p p l y v o l t a g e on the capacities is more serious. The valve capacities are not only dependent on the position of the electrodes in the valve, but also on the space charge between the electrodes. This space charge is in turn dependent on the anode current, which changes upon a variation of the supply voltage. Especially the capacity between cathode and control grid may hereby undergo considerable fluctuations which, without special precautions, would result in frequency variations.

The measures which may be taken to combat this phenomenon are of different kinds. The most conclusive means would be to stabilize the supply voltage. A suitable choice of connections for the oscillator valve may, however, help very much to keep the frequency variations mentioned small. We have used the following methods.

In the first place the influence of the valve capacity was in general limited by making the characteristic capacity C of the oscillation circuit as large as was possible in connection with the requirement that the oscillator must also be able to generate at short waves.

In order to make an oscillator generate it is necessary that the oscillation circuit should have a sufficiently high resonance resistance. The resonance resistance which can be obtained with an oscillation generally decreases with increasing capacity of the circuit. If therefore it is desired to obtain a sufficiently high resonance resistance at very high values of the resonance frequency, the capacity may not exceed a certain value. In order to make the value of the capacity at which oscillations can still be generated as high as possible, an oscillator valve with steep slope was used, namely the pentode EF 50[1]).

In the second place, as may be seen in fig. 1, the oscillation circuit was not placed in the grid circuit but in the anode circuit. Since the valve capacity between cathode and anode changes much less with the supply voltage than the capac-

[1]) The slope of the valve EF 50 is 6.5 mA/V. On the construction of this modern pentode see: A new principle of construction for radio valves, Philips techn. Rev. 4, 162, 1939.

ity between cathode and control grid, the frequency of the oscillation excited remains more constant with this arrangement. The influence of the changes of the grid-cathode capacity is not entirely eliminated in this way, since the capacity between control grid and cathode *via* the coupling coil M is transformed in the oscillation circuit in any case. In order to combat this effect the coupling between grid circuit and anode circuit is kept as loose as possible. Finally another simple measure is taken to combat the cause itself of the capacity variations, namely the fluctuations of the supply voltage. The rectifier for the supply is constructed for this purpose with a relatively high internal resistance. An increase in the supply voltage causes in general a fairly large increase in the plate current, and the internal voltage loss of the rectifier increases, so that a certain compensation of the voltage fluctuations is obtained. The result of all these measures is that the frequency variation with an oscillator frequency of 30 Mc/s for 10 per cent mains voltage change is not greater than about 6 000 c/s, *i.e.* 0.2 $^0/_{00}$. In the applications for which the apparatus is intended this variation may be entirely neglected. Any deviations between the frequency adjusted and that desired may then be ascribed more to the limitation of the accuracy of the scale than to frequency variations due to changes in the external conditions.

The regulation of the amplitude

In testing the sensitivity of a receiving set it is customary to apply to the aerial connections a high-frequency A.C. voltage such that a certain energy, 50 mW for instance, is fed to the loud speaker. The amplitude of the high-frequency signal then usually amounts to 1-50 μV in modern radio sets. The amplitude of the measuring oscillator must therefore also be able to be varied within this range. In the case of the apparatus under discussion, however, the voltage can be still further increased to 0.1 V. This is desirable when it is not a question of testing the set as a whole with the oscillator, but for instance only the intermediate-frequency stages which are designed for a considerably higher signal voltage.

The signal voltage given by the oscillator valve itself is of the order of magnitude of 10-100 V. If it is desired to make the oscillator generate stably, it is impossible to make the voltage much lower. In order to obtain the smallest desired signal, therefore, it is necessary to attenuate the voltage 10^7 to 10^8 times. An attenuation to 0.1 V is accomplished by coupling the "amplifier" stage following the oscillator stage only very loosely with the oscillator. By making this coupling loose enough not only is the desired reduction of the oscillator signal obtained, but at the same time, as already stated, there is the advantage that the influence of any changes in this amplifier stage or in the external loading is very much reduced.

Following the amplifier stage is an adjustable attenuator with which the signal voltage can be lowered continuously from 0.1 V to 1 μV (see *fig. 2*).

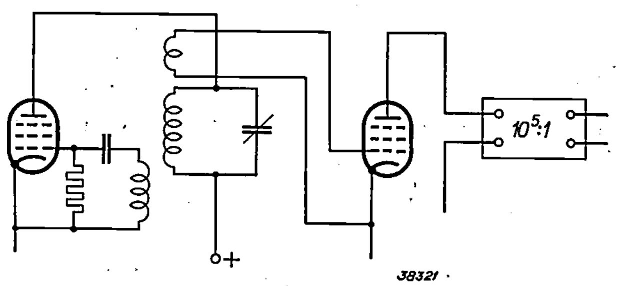


Fig. 2. Regulation of the amplitude of the oscillator voltage. The output voltage of the oscillator is applied to an amplifier stage *via* a very loose coupling. Following this is an attenuator which can regulate the voltage over a range 1 : 100 000.

In principle the attenuation of voltages is not difficult. An attenuator may be built up as a network of impedances, for which capacities, self-inductions or resistances may be chosen. In *fig. 3* an example is given of such an attenuator consisting of five cells, each of which attenuates the signal to $1/k$ of its original value. With $k = 10$, therefore, the desired attenuation by a factor of 100 000 will be obtained.

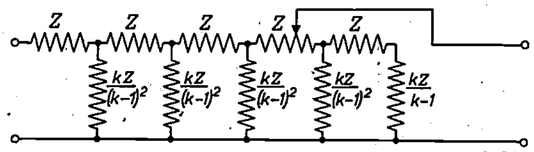


Fig. 3. Principle of an attenuator. In the network given each cell gives an attenuation of the input voltage by a factor k.

Since the attenuator must be able to function in a large frequency region, it is advantageous to use resistances as impedances, since resistances are in principle independent of the frequency. Practically, however, the branches of the attenuator will no longer behave exactly as resistances at the frequencies of short waves; the deviations become greater the higher the resistances. It is therefore advisable to choose small resistances. They must not, however, be so small that the self-induction of the necessary connection lines to the resistances becomes disturbing. A satisfactory compromise for the frequency region in question is formed by resistances between 10 and 300 ohms.

If the output signal must be very much attenuated by more than a factor 10^4 for, example, all kinds of difficulties occur at high frequencies. In that case various components are situated in the neighbourhood of the output terminals whose potential fluctuates many thousand times as much as that of the output terminals themselves. If by means of capacitative or inductive effects a small part of these potential fluctuations is transmitted to the output terminals, an interference voltage is obtained at the output of the measuring oscillator which may easily be much greater than the desired signal.

In order to avoid the capacitative and inductive effects mentioned it is necessary to keep the connections short, to shield very carefully the components which generate high-frequency electric fields and to take care that the shielding is properly earthed. For some other parts of the connections also, particularly for the attenuator itself and for the supply part, proper earthing is of the greatest importance for the avoidance of interferences. To illustrate this fact we shall consider in more detail only a detail of the attenuator. As may be seen in fig. 3 the lowest connection line of all the transverse resistances, which is earthed, also serves as common connection terminal of the input and output voltage. Let us now assume that the earthing of the connection line between the transverse resistances takes place at some other spot than that of the common connection terminal, so that between the connection line and the connection terminal there is an impedance z, although very small. In *fig. 4* this case is required, the electrical diagram shows what happens in this case.

It may be seen from the diagram that an interference voltage E_s acts on the output terminals in parallel with the desired output voltage. E_s is given by the following:

$$E_s = E_i \, z/R_i,$$

where R_i is the input resistance of the attenuator, which has a value of about 300 ohms.

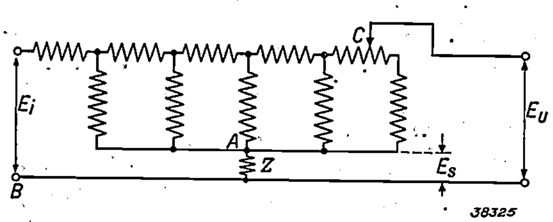


Fig. 4. Equivalent connections for the attenuator. Deviating from fig. 3, this diagram shows a certain impedance z between the connected transverse resistances of the attenuator and the corresponding connection terminal of the input and output voltage. This causes an interference voltage E_s which makes it difficult to obtain large attenuations.

If the input voltage must be able to be attenuated by a factor 10^5 without hindrance from this interference voltage, z/R_i must only be of the order of magnitude of 10^{-6}. z must therefore be smaller than 10^{-4} ohm. For a wave length of 5 m this means that the self-induction of the connection line must be smaller than 10^{-12} henry. A contact screw of 1 mm length may easily have a self-induction one hundred times as large. From this it will be sufficiently clear that in the construction of the attenuator very much care must be taken.

The impedance z in the attenuator, as already mentioned, is by no means the only source of interference voltages. Other interference voltages may be given, mainly by a capacitative coupling between different parts of the attenuator, furthermore by a capacitative coupling between the measuring oscillator and the apparatus to be tested, or by a coupling of these two apparatus *via* the power mains. Although all these effects can be combatted by the measures mentioned, their influence remains so strong that the attenuator may not be considered a precision instrument [2]). Since for certain applications, especially for selectivity measurement, it is advantageous to have at least one accurately known attenuation, a second attenuator which gives a fixed lowering of the amplitude in the ratio 1: 10 is introduced at the end of the connection cable which connects the measuring oscillator with the set being tested. This latter attenuator also serves as a socalled artificial aerial, *i.e.* it is built up of circuit elements of such kinds that the measuring oscillator seen from the output of the connection cable has about the same impedance as an ordinary aerial.

Securing the desired modulation

In sensitivity measurements on receiving sets the electrical energy which is delivered to the loud speaker is brought to a given value by regulation of the high-frequency voltage. Actually this energy is not determined by the amplitude, but by the modulation of the high-frequency signal, and therefore for a determination of the sensitivity it is necessary that the depth of modulation of the high-frequency signal should be known, while, moreover, it is desirable that the modulation should have a sinusoidal variation.

As to the frequency of the modulation, it is a

[2]) On the scale of the regulating knob of the attenuator voltage values are indicated. The actual output voltage may deviate from the voltage chosen as a result of the disturbing effects mentioned; the difference, however, amounts at the most to a factor 2. In this factor is included the drift of the output voltage with the frequency.

general custom to test the sets at a modulation frequency of 400 c/s. The reason for this is that this frequency is reproduced in full intensity by practically all sets, which is not the case for lower frequencies, while at higher frequencies some apparatus exhibit an irregular behaviour due to the occurrence of resonances. The modulation frequency of 400 c/s is generated by a separate oscillator included in the apparatus. Just as in the high-frequency oscillator, the resonance circuit is connected in the anode circuit. In order to obtain a truely sinusoidal signal inverse feedback is applied by means of a resistance in the cathode connection.

The A.C. voltage obtained in this way is now used to modulate the high-frequency signal, *i.e.* to influence it in such a way that there is a linear relation between the modulating voltage and the amplitude of the high-frequency signal. The method used here to accomplish this is so-called supressor grid modulation, whereby a negative bias voltage and in addition the modulating voltage are applied to the suppressor grid of a pentode. If there is a high-frequency A.C. voltage of constant amplitude on the control grid of a pentode, the amplitude of the anode A.C. is found to vary linearly with the modulating voltage within wide limits with a suitably chosen value of the suppressor grid bias voltage.

One of the problems which is encountered in the modulation of the oscillator signal is the occurrence of undesired frequency modulation. Suppressor grid modulation is accompanied by a periodic variation of the distribution of the space charge in the modulator valve (this is also true of every other method of modulation), and from this, as has already been mentioned, result periodic changes in the capacities between the electrodes. If the modulation were allowed to take place directly in the oscillator valve, which of itself is possible (see *fig. 5*), these capacity variations would lead to fluctuations of the oscillator frequency in the rhythm of the amplitude variations.

If such a signal which exhibits amplitude modulation as well as frequency modulation is now applied to the input of a selective high-frequency amplifier, the variation of the amplitude of the output signal may deviate very much from that of the input signal. The fluctuations of the frequency cause fluctuations in the output amplitude, especially when the frequency lies in the sloping part of the resonance curve. These amplitude variations are superposed on those which are caused by the amplitude modulation of the oscillator signal. The result may be a change in the depth of modulation, and at the same time a considerable distortion of the image of the modulation. *Fig. 6* gives two examples of this phenomenon: in case *a*) the

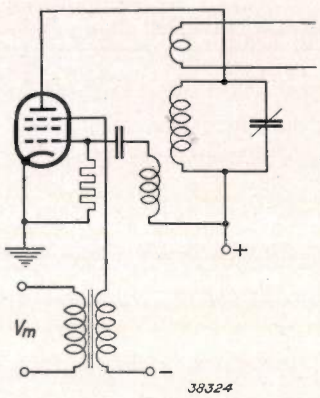
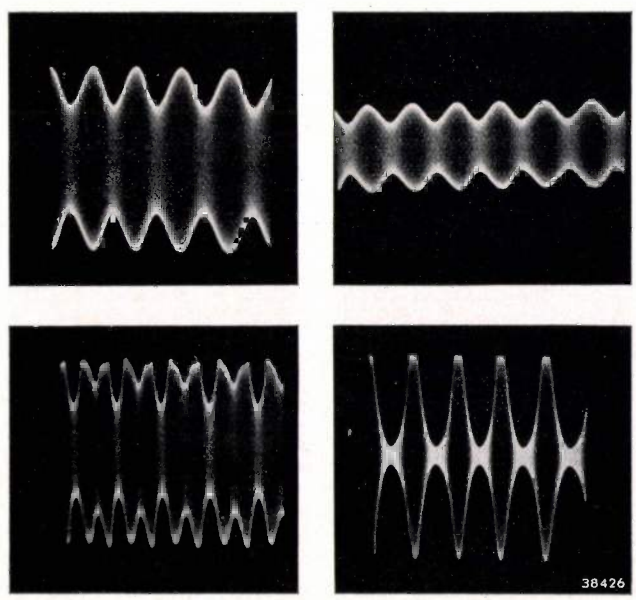


Fig. 5. Diagram for the modulation of the oscillator voltage with a signal of low frequency. These connections have the disadvantage that frequency modulation occurs.

average oscillator frequency corresponds to the resonance frequency of the receiving set, while in case *b*) the oscillator frequency lies on the sloping part of the resonance curve.

On the basis of the tuning curves of a large number of modern superheterodyne receivers a study was made of the degree of frequency modulation permissible when it is desired to be able to determine the sensitivity or the resonance curve of a



a *b*

Fig. 6. Examples of modulation distortion due to frequency modulation. A signal voltage which possesses amplitude modulation as well as frequency modulation is applied to an amplifier with a sharp resonance curve. In case *a*) the average frequency of the signal is equal to the resonance frequency, in case *b*) the signal frequency lies in the sloping part of the resonance curve. Above may be seen the input voltage, below the output voltage of the amplifier.

set with an accuracy of 10 per cent. It is found that for sensitivity measurements a frequency modulation greater than 200 c/s is not permissible, while for the measurement of the resonance curves of superheterodyne receivers this limit is only 30 c/s. But considering the fact that the selectivity of the receivers is determined practically only by the intermediate-frequency part, this last requirement only holds for relatively low frequencies.

How can the undesired frequency modulation be avoided? The modulation is made to take place not in the oscillator valve itself, but in the amplifier stage connected to it. Now the valve capacities of the amplifier valve do indeed change under the influence of the modulation, but since the amplifier stage, as already mentioned, is coupled only very loosely with the preceding oscillator stage, these variations have practically no influence on the resonance frequency of the oscillator circuit. In this way it is accomplished that the frequency modulation of 200 c/s permissible for sensitivity measurements is not exceeded in the whole range of oscillator frequencies up to 30 Mc/s. For frequencies up to 1.5 Mc/s (200 m) the frequency modulation even remains well within the limit of 30 c/s prescribed for selectivity measurements.

Construction of the apparatus

The larger the series in which an apparatus is manufactured, the greater the emphasis on easy methods of construction, while on the other hand the higher the permissible cost of the necessary tools. In this way in the manufacture of radio sets an extensive use of moulded parts has become prevalent. These parts necessitate expensive matrices, it is true, but they further simplify the manufacture very much.

Since the measuring oscillator is not manufactured in series of ten thousands, it was undesirable to make expensive matrices for the apparatus. For this reason no specially moulded parts were used. The chassis was built up as a closed body of sheet iron.

In the construction an attempt was made to collect the components as far as possible into small independent groups (units), which can be made and tested independently of each other. The wiring of the oscillator part was kept as short as possible, since the connection wires are always a source of individual differences between apparatus of a series, so that they might lead to difficulties in the calibration of the instruments.

A considerable shortening of the wiring of the oscillator circuit was obtained by a somewhat unusual construction of the wave length switch. The wave length switch is generally so constructed that it switches the connection terminals of the rotating condenser over to the different self-induction coils. For this purpose every coil must be connected to the wave length switch by means of a number of wires. In the measuring oscillator these connecting wires are avoided by using a rotating set of coils as wave length switch (see *fig.* 7). When the switch is given another turn, the

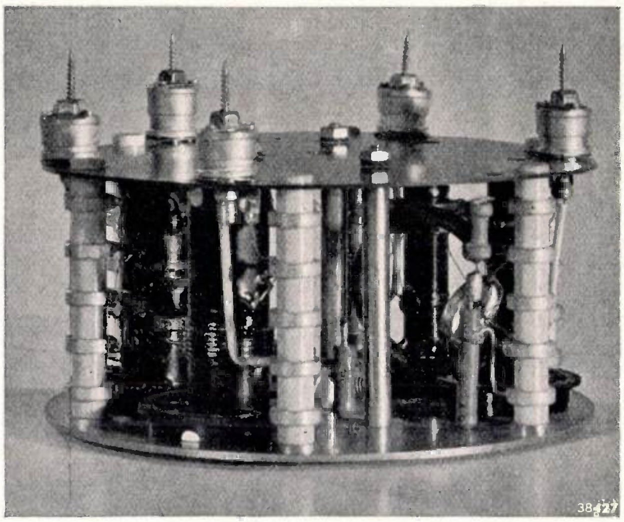


Fig. 7. The sets of coils of the measuring oscillator are mounted on the wave length switch so that upon switching over of the frequency region one set of coils (tuning coil and two coupling coils) is replaced by another set, while all connecting wires of the circuit remain in their places. For the region of shortest wave lengths an oscillator coil of one winding is used, which may clearly be seen in the figure; the remaining coils are wound on coil bobbins placed on the base plate.

self-induction coil and the corresponding back-coupling coil are replaced by another combination of coils, while otherwise all the connections and wiring remain unchanged. The capacities and self-inductions of the wiring thus also remain unchanged and may be kept very small, so that at every position of the rotating condenser one is certain that the wave lengths of the oscillator signals which are successively excited upon rotating the wave length switch have exactly the same relation as the square roots of the self-inductions of the coils. This has the advantage that for different frequency regions, such as

0.1— 0.3 Mc/s
1.0— 3.0 Mc/s
10 —30 Mc/s

the same scale can be used.

Fig. 8 shows the rear side of the apparatus opened, from which the arrangement of the main components may be seen. In order to avoid a heating up of the oscillator circuit as far as possible, the supply part and all the valves are shielded with respect to the rotating condenser and the set of coils. Moreover, the latter are placed at the bottom, while the

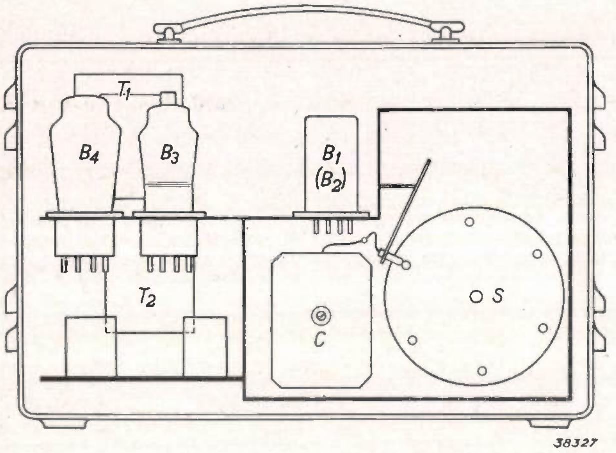


Fig. 8. Rear view of the opened measuring oscillator GM 2 882. B_1 oscillator valve shielded by a can, B_2 the valve directly behind B_1 of the "amplifier" stage, B_3 oscillator valve for the modulation frequency, B_4 rectifier valve of the supply apparatus; T_1 supply transformer, T_2 transformer of the modulator stage, C tuning condenser, S rotating set of coils. In the arrangement of the components the principle was followed that the components which develop the most heat should be placed at the top and separated from the condenser and set of coils by partitions.

sources of heat are at the top and can give off their heat by means of air circulation.

The rotating condenser is driven by means of a friction coupling *via* a second coupling which is flexible in an axial direction, so that external forces on the tuning knob can have no effect on the distance between the plates of the two sets. The pointer of the scale is driven by means of a string transmission by the condenser axis. By adjustment of the rotating condenser provision is made in every apparatus that the scale gives exactly the correct value at seven points: the possible errors at other points on the scale then remain within the limits of error in reading off. Scale and pointer are behind a glass window, so that damage during transportation is impossible. Together with the flexible transmission between the tuning knob and the condenser this offers the greatest guarantee that the calibration will remain accurate for a long time.

Fig. 9 finally gives a view of the outside of the measuring oscillator. In addition to the possibilities of adjustment already discussed, namely the switch for the wave ranges, the rotating condenser and the attenuator, there is only one other operating knob. This is a combination switch for switching the apparatus on and off and for passing from internal modulation to external modulation. In the latter case the desired modulation voltage can be applied from the outside; the valve of the low-frequency oscillator then serves as an amplifier valve.

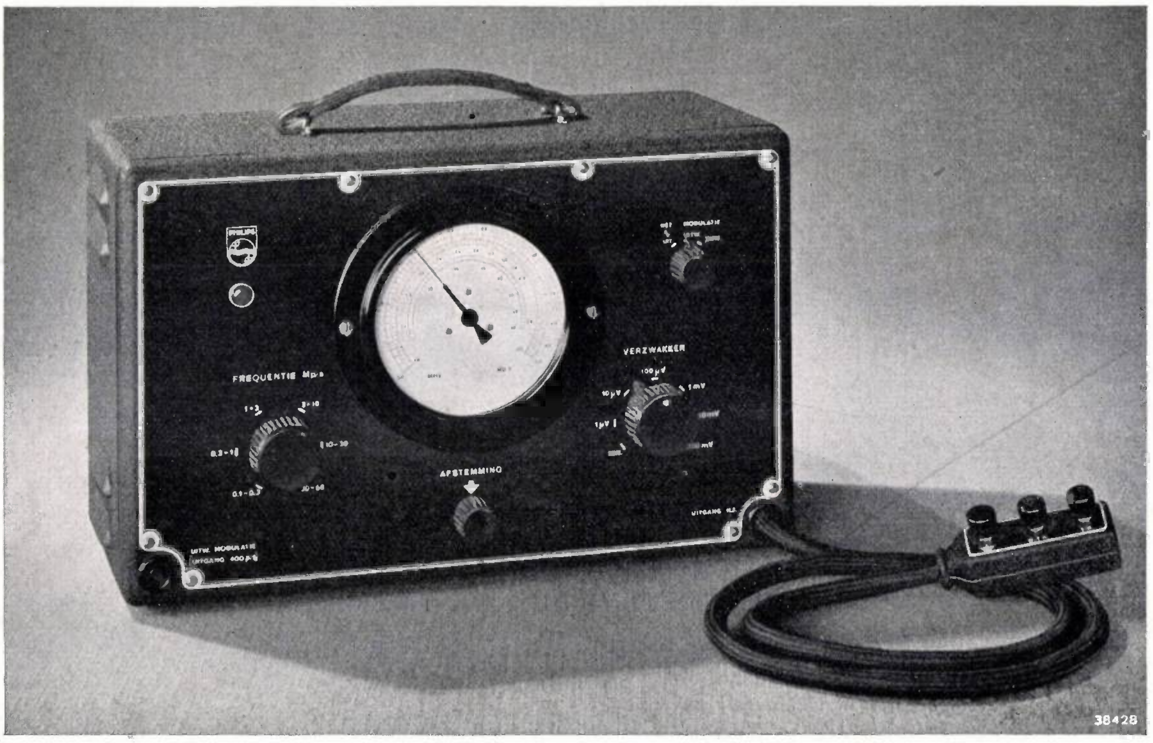


Fig. 9. View of the outside of the measuring oscillator GM 2 882. The apparatus has only four operating knobs, namely the tuning knob, the wave switch, the attenuator and a combination switch for switching the apparatus on and off and for choosing internal or external modulation.

VOL. 6. No. 6, p. 161-192 JUNE 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

THE CONCEPT OF BRIGHTNESS IN CONNECTION WITH BLACKOUT PROBLEMS

by P. J. BOUMA.

535.241.44 : 628.978.7

The classical concept of brightness determined with the help of the standard eye-sensitivity curve leads to difficulties in cases where one is concerned simultaneously with low levels of brightness and different coloured light sources (road lighting, blackout problems). The different possibilities to be considered in introducing a more suitable concept of brightness are discussed, and the determination of the quantity concerned is divided into a definition of equality (when may two brightnesses be considered equal?) and a definition of degree (what value must be assigned to a given brightness?). The mathematical formulation only becomes simple when we assume the validity of the summation law for estimating the brightness: the brightness then becomes an arbitrary function of $S(\lambda)\,V(\lambda)\,d\lambda$, where $S(\lambda)$ represents the spectral distribution of energy and $V(\lambda)$ the eye-sensitivity curve under the given conditions of observation. By the choice of different values of this function the definitions of degree of different concepts of brightness are obtained, such as Purkyne brightness, "Dunkelleuchtdichte", etc. The validity or non-validity of the summation law under various conditions of observation is discussed in detail. In conclusion observability of coloured point sources of light is discussed. For the calculation of the threshold value of visibility it is found that it is not permissible to start with the eye-sensitivity curve for rod vision, and that no summation law holds. From this it follows that the so-called "Dunkelbeleuchtungsstärke" which occurs at the eye is not a measure independant of colour for the observability of a light point.

In a previous article published in this periodical [1]) it was explained how it was desirable in the study of the problems of modern road lighting, where it is a question of relatively low levels of brightness and very different colours, to introduce in addition to the customary concept of brightness a new concept which is more closely related with our intuitive concept of "brightness" under the given conditions.

In connection with the problems of blackout, where still much lower brightnesses occur, and where the unsuitability of the ordinary idea of brightness is particularly striking, this problem has recently attracted considerable attention.

In the following we shall therefore consider the problem of the determination of the concept of brightness again, and this time from a general point of view. After having explained the requirements which every definition of brightness must satisfy, we shall study more closely the different possibilities of arriving at a practically useful determination.

The definition of equality

When the eye observes two luminous planes adjacent to each other, it is very difficult to estimate the brightness relation between the two planes; on the other hand the eye can indeed judge which plane is the brighter, and we can adjust the brightness of one of the planes with considerable accuracy so that the two planes give the impression of being "equally bright".

In setting up a definition of brightness we must keep this property of the eye in mind, so that the first condition which the definition must satisfy is the following.

When two planes give to the eye the impression of being "equally bright" the newly defined brightness must have the same value for both cases [2]).

This requirement may be considered as the "definition of equality": by definition we have determined under what circumstances we shall consider two brightnesses equal. This definition will of course only have a practical value when by means of an extensive experimental investigation we have ascertained the quantities of light of different colour which must be emitted under various conditions

[1]) Philips techn. Rev. 1, 142, 1936.

[2]) It is clear that this requirement can only be satisfied when there is transitivity of the impressions of brightness, in other words: when two planes A and B give the impression of being "equally bright" and the same is true of B and C, then A and C upon direct comparison also give the impression of being equally bright. This transitive law is found to be valid under all conditions.

to give the average observer the impression of equal brightness. This investigation has by no means been completed for all conditions occurring in practice.

The definition of equality is already sufficient for the treatment of many problems. For certain problems, however, it will be a great advantage to add a "definition of degree" which enables us to assign to a brightness a definite value. The definition of equality answers the question "When are two brightnesses equal?", the definition of degree the question "How great is the brightness?"

The definition of degree in general

When we consider how to introduce a definition of degree it is immediately clear that we may do this arbitrarily to a certain extent. This fact is connected with the fact that the eye is not able to estimate directly the ratio of the brightnesses of two planes. In the choice of the definition of degree we are, however, bound by two conditions:

1) Care must be taken not to come into conflict with the definition of equality.
2) When one of two planes is found to be brighter that the other, the brightness of the first must be assigned a higher value.

The most obvious method, and therefore the one appearing earliest in the literature (König 1891), is the following: an arbitrary kind of comparison light is chosen, and by definition it is established that for that kind of light the brightness increases proportionnally with the energy radiated per unit of surface. This establishes for this one kind of light a complete scale of brightnesses, while for all other kinds of light the value of the brightness can be determined by means of the definition of equality [3]). It is clear that in this way the two conditions mentioned above are automatically fulfilled.

The kind of comparison light spoken of may be chosen quite arbitrarily, and different formulations of the concept of brightness occurring in the literature are distinguished from each other only in the choice of the comparison light. We shall give a few examples.

König used as comparison light a spectral colour with the wave length 5 350 Å. This colour was also used by the present author in defining the concept of "subjective brightness".

The concept of "Dunkelleuchtdichte" which has been introduced in Germany is determined in quite the same way, but as comparison light the radiation of a black body with a temperauter of 2 360° K [5]) has been chosen. Such a choice has the advantage that with a visual photometer, containing as comparison lamp an electric lamp of this colour temperature, the "Dunkelleuchtdichte" can be correctly measured directly (provided the eye of the observer is normal and well adapted). Since the concept of "Dunkelleuchtdichte" was introduced especially for blackout problems, in working out a method of calculation for the "Dunkelleuchtdichte" the investigation could be limited to those brightnesses at which pure rod vision occurs. For greater brightnesses this concept may not be used.

Quite a different example of a definition of degree would be to choose the brightness scale for a single definite kind of light so that with that kind of light the logarithm of each brightness ratio $B_1 : B_2$ is a direct measure of the number of steps lying between B_1 and B_2 that have a difference in brightness which is just observable (law of Weber and Fechner). For other kinds of light also this law would then be approximately valid. Over against this, however, are two disadvantages, namely the awkwardness of the concept for the purposes of calculation and the fact the definition does not pass over at high brightnesses into the ordinary brightness definition, as is the case for the other two definitions of degree.

Eye-sensitivity curve and summation law

In the further development of the definition of brightness the ideas of "eye-sensitivity curve" and "summation law" play an important part. An eye-sensitivity curve is obtained by determining for a given level of brightness the energy (in watts/cm^2 for instance) which must be emitted in the form of different monochromatic radiations to give the eye the impression of "equal brightness". When this energy has the value $s(\lambda)$ for a given wave length λ, while $s(\lambda_0)$ is the value for the fixed comparison wave length λ_0 for which the energy in question is a minimum, the eye-sensitivity curve is the function $V(\lambda) = s(\lambda_0) : s(\lambda)$. For $\lambda = \lambda_0$, $V(\lambda)$ assumes the value of unity for all other wave lengths $V(\lambda) < 1$. The shape of the function $V(\lambda)$ depends very closely upon the conditions under which the comparison takes place (level of brightness, size of the field of vision, method of photometry, etc.). These conditions must be kept constant during the recording of an eye-sensitivity curve and they must of course also be taken into account in the application of the eye-sensitivity curve.

---

3) In the corresponding acoustic case (the determination of the concept of loudness) the same method is used: as comparison sound a tone is chosen of 1 000 c/s.

4) Philips techn. Rev. 1, 142, 1936.

5) A. Dresler, Das Licht 10, 112, 118, 145, 1940.

By the summation law we mean the following property of brightness impressions.

When a radiation of energy S_1 (spectral distribution 1) and a radiation of energy S_2 (spectral distribution 2) give the same impression of brightness, this is also true of the mixture $aS_1 + (1-a) S_2$, where a is any arbitrary number between 0 and 1.

The validity of the summation law

A detailed investigation of the validity of the summation law at low brightnesses and with fairly large fields of vision was carried out by the writer [6]). The following colours were chosen: the light of incandescent filament lamps, fairly saturated red, blue and green (filtered electric light) and the mixed colours formed from these by mixing two colours in arbitrary proportions. All the colours were compared directly with electric light. For each mixture the brightness to be expected from the summation law was determined and compared with that measured. The range of brightnesses extended from 5×10^{-5} to 0.3 c.p./m^2, and thus included the greater part of the region important in road lighting and the whole region with which we are concerned in blackout problems. With fields of vision of at least several degrees the summation law was found to hold very accurately in this range of brightnesses. In the comparison between calculated and measured values the average deviations amounted to about 1 per cent and exhibited quite the character of accidental errors. Only at the highest brightnesses used were the measurements more or less non-reproducible in the neighbourhood of the saturated blue, and there were indications that deviations from the summation law began to occur here.

Much work has already been done on the question of the validity of the summation law in the region of high brightnesses (pure cone vision), but without reaching consistent results on this point, which is of extremely great importance in photometry [7]). The main cause of this lack of agreement is that in this region of brightness, where the colour difference is not weakened by the collaboration of the rods, the "natural" method of measuring the brightness, namely the direct comparison of two different coloured planes side by side, is extremely difficult and leads to very irreproducible results. Some authors (Helmholtz, v. Kries) even doubt whether such measurements have any value at all. For this reason attempts were soon made to replace this unsatisfactory method of photometry by other methods which must satisfy the requirements of accuracy and reproducibility, of not deviating too much in their results from those of direct comparison and of obeying the summation law. After several vain attempts two methods were found which not only satisfied the requirements made of them, but which also gave good mutual agreement in their results. These were the flicker method [8]) and the step-by-step method [9]). In 1924 it was therefore decided to standardize a mean of the best of these measurements carried out by both the methods as international eye-sensitivity curve $V_k(\lambda)$, and to base the concept of brightness for high levels of brightness upon this curve. Since then the criterion for the excellence of a photometric method is the question of whether or not it produces results which agree with this standardized definition within the limits of accuracy required for the purpose in view.

Although this created a very satisfactory situation for practical work, numerous investigations were still carried out on the mutual agreement between the different photometric methods and on the validity of the summation law. The most important results are the following:

1) For the flicker photometer the summation law is very accurately valid (to within 1 to 2 per cent [10])).
2) The step-by-step method gives results which agree within several per cent with those from the flicker photometer [9c]), and here also the summation law is valid. Although numerous investigators confirmed these two facts, they were pertinently contradicted by others [11]).
3) The method of direct comparison does not lead to reproducible results. According to certain authors [12]) there is good agreement with the flicker photometer, according to others there are fairly large, but unsystematic deviations [8b]), others again find systematic deviations of 10-20 per cent [8c]), or even of 50-100 per cent [11, 13]). The last authors find the greatest deviations for

6) Proc. Kon. Akad. Wet. Amsterdam 38, 150, 1935.

7) *Cf.* for this part also Philips techn. Rev., 1, 120, 1936 and 5, 283, 1940.

8a) H. E. Ives, Phil. Mag., 24, 149, 1912.
b) W. W. Coblentz and W. B. Emerson, Bull. Bur. Stand. 14, 167, 1918.

9a) H. E. Ives, Phil. Mag. 24, 744, 1912.
b) E. P. Hyde, W. E. Forsythe and F. E. Cady, Astrophys. J. 48, 87, 1918.
c) K. S. Gibson and E. P. T. Tyndall, Sc. Papers, Bur. Stand. 19, 131, 1923.

10) The small deviations (2%) found by Jaggi may probably be ascribed for a large part to the lack of absolute constancy of the eye-sensitivity curve.

11) A. Kohlrausch, Das Licht 5, 259, 275, 1935.

12) R. G. Weigel, Das Licht 5, 43, 1935.

13) A. Dresler, Das Licht 7, 81, 107, 1937.

saturated red, blue-green and blue, the smallest for yellow.

The cause of these discrepancies is not yet completely explained. Many factors which are difficult to control play an important part: technique, and speed of adjustment, the experience of the observer, the psychological attitude, the ability to abstract the colour difference, consciously or unconsciously applied devices of accomplishing this latter, etc.

In the cases in which the above-mentioned large deviations from the results obtained with the flicker photometer were found, the summation law also fails to hold: too high brightnesses are assigned to the saturated colours at the extremities of the spectrum, and this phenomenon disappears when these colours are mixed to give an unsaturated colour. Thus lower values of the brightness are found for the mixture than would be expected according to the summation law.

Summation law and definitions of degree

If the validity of the summation law is assumed the following property can be derived from the definition of the eye-sensitivity curve $V(\lambda)$.

Two planes which emit a radiation with the spectral distribution $S_1(\lambda)$ and $S_2(\lambda)$ will give the impression of being "equally bright" when

$$\int S_1(\lambda)\, V(\lambda)\, d\lambda = \int S_2(\lambda)\, V(\lambda)\, d\lambda,$$

where $V(\lambda)$ represents the eye-sensitivity curve that is valid under the conditions at which the comparison takes place. $S(\lambda)\, d\lambda$ is the energy which is emitted in the wave-length region between λ and $\lambda + d\lambda$ per cm² (projected in the direction of observation), per second and per unit of solid angle (steradian).

From the definition of $s(\lambda)$ as an intensity which stimulates the same impression of brightness for all wave lengths, it follows directly with the help of the summation law that the mixture $a_1 s(\lambda_1) + a_2 s(\lambda)$ gives the same impression of brightness as $s(\lambda_1)$ or $s(\lambda_2)$ of $a_1 + a_2 = 1$. If this property is extended to more components, and if one then passes over to a continuous spectrum by transition of the limits, it is found that the mixture $\int a(\lambda)\, a(\lambda)\, d\lambda$ makes the same impression of brightness as $s(\lambda)$ if $\int a(\lambda)\, d\lambda = 1$. If $a(\lambda)\, s(\lambda)$ is set equal to $S(\lambda)$, it follows that for two spectral distributions $S_1(\lambda)$ and $S_2(\lambda)$ which give the same impression of brightness the following is valid:

$$\int \frac{S_1(\lambda)}{s(\lambda)}\, d\lambda = \int \frac{S_2(\lambda)}{s(\lambda)}\, d\lambda$$

or, introducing the eye-sensitivity curve $V(\lambda) = s(\lambda_3)/s(\lambda)$

$$\int S_1(\lambda)\, V(\lambda)\, d\lambda = \int S_2(\lambda)\, V(\lambda)\, d\lambda.$$

This relation enables us to arrive at a common mathematical formulation of all the possible definitions of degree. The new brightness which is to be introduced is connected, according to the above, unambiguously with the expression $\int S(\lambda)\, V(\lambda)\, d\lambda$, so that we find directly as the most general mathematical formulation of the definition of degree:

$$B = f \left\{ \int S(\lambda)\, V(\lambda)\, d\lambda \right\} \quad . \ . \ . \ . \quad (1)$$

In this equation f represents an arbitrary function of which it is only required that upon continuous growth of the integral, B shall also increase continuously. Formula (1) is of course only valid in conditions under which the summation law is valid[14]). For great brightnesses this is the case, as stated, only if the flicker photometer is used as a measuring instrument. Since at the same time one also finds very approximately the same eye-sensitivity curve $V_k(\lambda)$, the simplest definition of degree for great brightnesses is

$$B_c = C_k \int S(\lambda)\, V_k(\lambda)\, d\lambda, \quad . \ . \ . \ . \quad (2)$$

where $V_k(\lambda)$ is the internationally established standard eye-sensitivity curve and C_c is a constant which takes on the value 636 when the brightness B_c is expressed in stilb (c.p./cm²) and $S(\lambda)$ in watts/cm² [15]). This formulation is internationally established and the concept of brightness thus defined will in the following be called the "classical brightness" and indicated by the letter B_c.

When we do not confine ourselves to great brightnesses, the simplest relation between brightness and the integral is also:

$$B = C \int S(\lambda)\, V(\lambda)\, d\lambda, \quad . \ . \ . \ . \ . \quad (3)$$

a definition which was proposed by Voet-Mogendorff[16]) among others under the name of the Purkyne brightness. If C is chosen equal to C_c, then for the conditions under which the standard eye-sensitivity curve is measured, equation (3) passes over into the classical brightness B_c according to equation (2).

The definitions with a comparison light discussed

14) In the acoustic case for example the summation law does not hold, so that such a simple mathematical formulation is impossible there.

15) The letters H and E used in the earlier article [1]) have been replaced by the modern notation B and S, while a subscript c is added to indicate that we are here concerned with the classical concept of brightness and with the standard eye-sensitivity curve. Since equation (2) indicates the classical brightness of a source of radiation as a function of purely physical properties, it is possible to measure classical brightnesses without using the eye as an instrument. This has led to the regular use of the concept of classical brightness even under conditions where the curve $V_k(\lambda)$ is no longer valid. Under these conditions the classical concept of brightness will not satisfy the definition of equality.

16) H. H. Voet-Mogendorff, Diss. Amsterdam 1939.

above also follow from (1) by means of a definite choice of the function f. For the comparison light B is everywhere proportional to the energy radiated and the new concept of brightness coincides with the classical. From this it follows that for ranges of brightness in which $V(\lambda)$ does not change with the brightness, B must always increase proportionally with the integral. If a field of vision of several degrees is chosen, this proportionality occurs in the region of pure cone vision (great brightnesses) and the region of pure rod vision (very low brightnesses).

In the first case (1) passes over into the classical definition (2); while in the second case

$$B_s = C_s \int S(\lambda)\, V_s(\lambda)\, \mathrm{d}\lambda, \quad \dots \quad (4)$$

If we confine ourselves to these two limiting cases we can calculate directly from the classical brightness B_c the new brightness (subjective brightness, Purkyne brightness, "Dunkelleuchtdichte"), by dividing formulae (2) and (4) by each other:

$$\frac{B_s}{B_c} = \frac{C_s}{C_c} \frac{\int S(\lambda)\, V_s(\lambda)\, \mathrm{d}\lambda}{\int S(\lambda)\, V_k(\lambda)\, \mathrm{d}\lambda} \quad \dots \quad (5)$$

The ratio C_s/C_c is found by substituting in (5) the spectral distribution $S_0(\lambda)$ of the comparison light, for which B_s must equal B_c. We find the general formula:

$$\frac{C_s}{C_c} = \frac{\int S_0(\lambda)\, V_k(\lambda)\, \mathrm{d}\lambda}{\int S_0(\lambda)\, V_s(\lambda)\, \mathrm{d}\lambda} \quad \dots \quad (6)$$

If we choose for the comparison light a spectral colour with the wavelength λ_0 then (6) passes over into:

$$\frac{C_s}{C_c} = \frac{V_k(\lambda_0)}{V_s(\lambda_0)} \quad \dots \dots \quad (7)$$

From this formula it may be seen especially

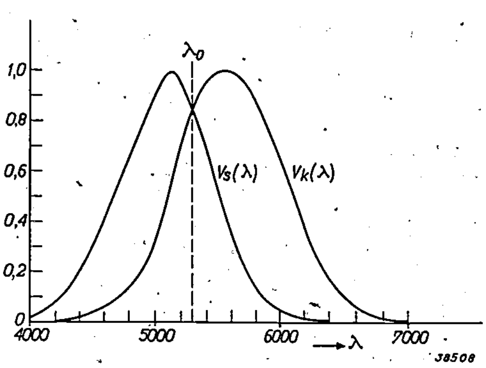


Fig. 1. Eye-sensitivity curves for fields of vision of several degrees in diameter. V_k holds for great brightnesses (standard eye-sensitivity curve), V_s holds for very low brightnesses (region of pure rod vision). The two curves intersect at λ_0 = 5 290 Å.

that $C_s = C_c$ when as comparison light the spectral colour is chosen of the wave length λ_4 = 5 290 Å, where the curves $V_k(\lambda)$ and $V_s(\lambda)$ intersect each other (see *fig. 1*). For the different kinds of comparison light mentioned above formulae (6) and (7) give the following results:

Name of the concept of brightness	Comparison light	C_s/C_c
Subjective brightness	λ_0 = 5 350 Å	1.20
"Dunkelleuchtdichte"	black radiation 2 360 °K	2.18
Definition with λ_0 = 5 290 Å	λ_0 = 5 290 Å	1.00

Fig. 2 shows schematically the shape of the function f. The brightness B and the expression $\int S(\lambda)\, V(\lambda)\, \mathrm{d}\lambda$ are here considered to be plotted in the same logarithmic scale. The Purkyne brightness (*1*) then gives an entirely linear relation.

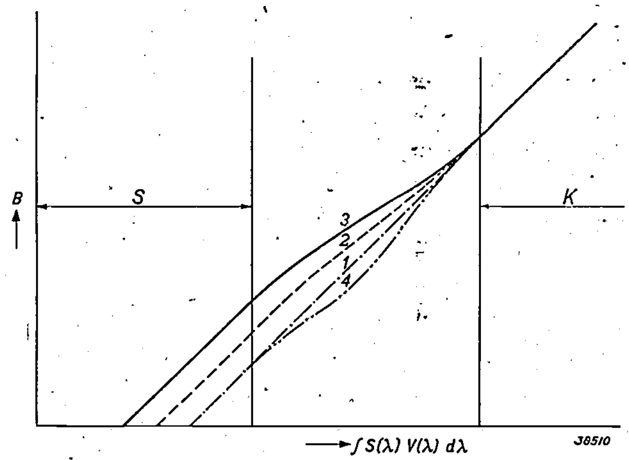


Fig. 2. Graphical representation of the relation given in formula (1) between the brightness B and the expression $\int S(\lambda)\, V(\lambda)\, \mathrm{d}\lambda$.
1 Purkyne brightness (straight line).
2 Subjective brightness } (two straight sections not in the same straight line joined
3 "Dunkelleuchtdichte" } by a curve).
4 Brightness definition with λ_0 = 5 290 Å as comparison light (two straight sections in the same line).
S gives the brightness range within which there is pure rod vision, K the region in which pure cone vision occurs. B and the integral are considered to be plotted on a logarithmic scale.

The subjective brightness (*2*), the "Dunkelleuchtdichte" (*3*) and the concept of brightness based on the comparison wave length λ_0 = 5 290 Å (*4*) give two straight sections, joined by a curved line. Only the left-hand section of the "Dunkelleuchtdichte" may be used; in the case of curve (*4*) the two straight sections are an extension of each other.

It is uncessary to find new names for the units of the new concepts; particularly for the definitions with comparison lights it is quite possible to say, for instance "the classical brightness of a plane is

3×10^{-4} c.p./cm², the subjective brightness only 2×10^{-4} c.p./cm²", which expresses the fact that the plane appears as bright as one which radiates monochromatic light of 5 350 Å and possesses a brightness of 2×10^{-4} c.p./cm². For practical reasons, however, a new name has been introduced in Germany for the unit of "Dunkelleuchtdichte", namely the "Skot" instead of $10^{-7}\pi^{-1}$ stilb.

Visibility of point sources of light

For light spots observed at an angle of several degrees the summation law and the eye-sensitivity curve $V_s(\lambda)$ are found to retain their validity even at extremely low brightnesses in the neighbourhood of the threshold of the brightness observation, and the subjective brightness (or the "Dunkelleucht-dichte") at which the threshold value is reached is the same for all colours.

In blackout problems, however, one is concerned not only with the threshold value for large spots of light, but even more with the limits of visibility for light spots of extremely small dimensions. When the spots are sufficiently small their cross section is of no concern, and the visibility is determined exclusively by the amount and the colour of the light cast on the eye of the observer [17]), *i.e.* of the intensity of illumination E on the eye of the observer. For this concept ("Punkthelle", "éclat apparente") we shall use the term "eye illumination". If by E_c one means the value of the classical concept of illumination intensity (based upon the standard eye-sensitivity curve $V_k(\lambda)$), it is clear that the threshold value E_c will not be the same for all colours. By analogy with formula (5) which shows how it is possible to calculate the value B_s of the new concept of brightness (for instance, the "Dunkelleuchtdichte") which has the same threshold value for all colours for large light spots from the classical brightness B_c, it seems obvious to assume that from the classical illumination intensity E_c, with the help of an analogous formula

$$\frac{E_s}{E_c} = \frac{C_s}{C_c} \frac{\int S(\lambda)\, V_s(\lambda)\, \mathrm{d}\lambda}{\int S(\lambda)\, V_k(\lambda)\, \mathrm{d}\lambda} \quad . \quad . \quad . \quad (8)$$

we should be able to calculate a "Dunkelbeleuchtungsstärke" at the eye which would then have to have the same value for all colours at the threshold of visibility.

In the modern German literature this assumption is accepted as obviously correct. For the unit of the new concept of intensity of illumination the name "Nox" instead of millilux is used.

[17]) Philips techn. Rev., 4, 15, 1939.

On the basis of certain experimental data we shall now test the above assumption about the visibility of light points. In *table I* for six different kinds of light, all of which fall within a fairly narrow spectral region, the ratio is given between the threshold value of the eye illumination for coloured light (E') and ordinary electric light (E).

Table I

Colour	Authors	I	II	III
		$E_c' : E_c$	$E_s' : E_s$	$\overline{E_s'} : \overline{E_s}$
red	Langmuir	16,8	0,80	1,75 (0%)
yellow (Na)	Weigel	4,1	0,83⁵	1,56 (—11%)
yellow (Na)	Bouma	5,3	1,08	2,02 (+15%)
green	Langmuir	1,00	1,59	1,70 (—3%)
blue	Langmuir	0,310	4,25	1,92 (+10%)
blue	Arndt	0,258	3,44	1,54 (—12%)

Column I gives the ratio of the classical illumination intensities $E_c' : E_c$. As was to be expected this ratio depends very much upon the colour chosen. Column II gives the ratio between the "Dunkelbeleuchtungsstärken" calculated according to formula (8) $E_s' : E_s$. According to the assumption made the value 1.00 should be obtained for all colours. It is clear, however, that there are large deviations, especially in the blue, from which it is evident that the assumption referred to is quite incorrect.

Since the observation of points of light is quite a different problem from the observation of large spots of light, this is not surprizing, and the question arises whether better agreement could be obtained be replacing the function $V_s(\lambda)$ in (8) by another function $\overline{V_s}(\lambda)$, which is better adapted to the altered conditions of observation. In column III the ratios $\overline{E_s'} : \overline{E_s}$ are given which are found when $\overline{V_s}(\lambda)$ is so chosen that the threshold values become as nearly as possible equal for the different coloured lights.

The striking result is obtained that the ratio $\overline{E_s'} : \overline{E_s}$ becomes reasonably constant, but that its average value is greater than unity, namely 1.75 (the deviations from this average value are given in parenthesis in the table). This means that for point sources of light the following assumption analogous to the summation law cannot be correct:

"When a radiation with the enrgy S_1 (spectral distribution 1) and a radiation with the energy S_2 (spectral distribution 2) both represent a threshold value, this is also true of the case of the mixture $aS_1 + (1-a) S_2$".

Practical conclusions and applications to blackout problems

We shall now give a few applications of the above results to blackout problems.

a) The use of the concepts "Dunkelleuchtdichte" and "Dunkelbeleuchtungsstärke"

From the above the conclusion may be drawn that for the road user who observes a large field of vision at low illumination intensities the brightness measured in "Skot" is a good measure of the visibility, but that for an air pilot, who will usually have to observe objects in the shape of points, the visibility is certainly not connected unambiguously with the "Nox" value of the eye illumination. In the case of light spots of at least several degrees in diameter, therefore, the introduction of the new concepts of brightness facilitates the calculation of the threshold value, but this is not true in the case of light points. If for example one considers the last colour mentioned in table I, then when the classical eye illumination (column I) is used it must be stated that the threshold value expressed in lux for white light is about 3.9 times that for blue light, while when the "Dunkelbeleuchtungsstärke" according to column II is used, it must be stated that the threshold value expressed in "Nox" for blue light is about 3.4 times that for white light. It is clear that in this case the introduction of the new concepts has brought no advantage at all.

The deviations observed in column II of table I may also be formulated in the following way: for the air pilot who observes points of light the threshold value for blue light lies considerably higher than would be expected, or, for blue light the Purkyne effect offers the air pilot much smaller advantages than the road user. Such a difference may be considered as an advantage in the use of blue light, to which statement it must immediately be added that this holds only for the case in which the blue light is used exclusively to make possible the general orientation of the road user. In all cases where it is a question of the recognition by the road user of definite signs (for instance letters as an indication of shelters) blue light has important disadvantages, since visual acuity is very low with this light.

b) Estimation of the threshold value for coloured point sources of light

In order to reach an estimation of the threshold value for coloured point sources of light in spite of the proven unsuitability of the concept of "Dunkelbeleuchtungsstärke" and the non-validity of the summation law, the experimental data of table I have been put into the form shown in *fig. 3*. In the horizontal direction the ratio of the threshold values of the classical eye illumination (E_c' for coloured light, E_c for ordinary electric light) has been plotted, as it would be calculated on the assumption that the "Dunkelbeleuchtungsstärke" or the subjective brightness was here a good measure of the visibility, in other words on the assumption that $E_s' : E_s$ were equal to unity.

In the vertical direction the experimental values of $E_c' : E_c$ are plotted. The values found are joined by the full drawn line. The figure is valid only for light sources which emit the greatest part of their light in a relatively narrow spectral region. In order to determine the threshold value of such a light source, $(E_c'/E_c)_{\text{theor.}}$ is first calculated according to the formula:

$$E_c' : E_c = \frac{E_s}{E_c} : \frac{E_s'}{E_c'},$$

where the two quotients on the right-hand side can be determined with the help of equation (8). Then with the help of fig. 3, one finds the corresponding value of $(E_c'/E_c)_{\text{exp.}}$.

For light sources whose radiation is distributed over the whole spectrum fig.3 is not valid, and no conclusions can be drawn. An experimental determination will be necessary here. For small spots of light, which may not yet be considered as point sources (between about 2′ and 1°), the full drawn curve of fig. 3 must be replaced by another curve which passes between the full line and the dotted straight line, and is closer to the latter the larger the light spot.

c) The use of the filter method for blackout purposes

The filter method consists in the combination of sources of coloured light and coloured glass window panes such that while the window pane transmits a large percentage of daylight it absorbs almost all of the coloured artificial light [18]). When this method is used the small percentage of light which reaches the outside is usually of a pronounced colour, so that care must be taken in calculating the threshold value.

If sodium light is combined with green window panes, the light emitted contains practically only the yellow sodium lines. From table I it may be seen that for this colour the threshold value of the eye illumination lies about 4.7 times as high as

[18]) Philips techn. Rev., 5, 93, 1940.

that for ordinary electric light. This factor cannot, however, be completely taken into account, since sodium light has such a striking colour that the correction factor N which must be introduced in

If electric lamps in orange-yellow glass bulbs are combined with green windows, it may be calculated from the transmission curve of Matthews and van Liempt [18]) (fig. 2) that $(E_c' : E_c)_{\text{theor.}} = 0.62$.

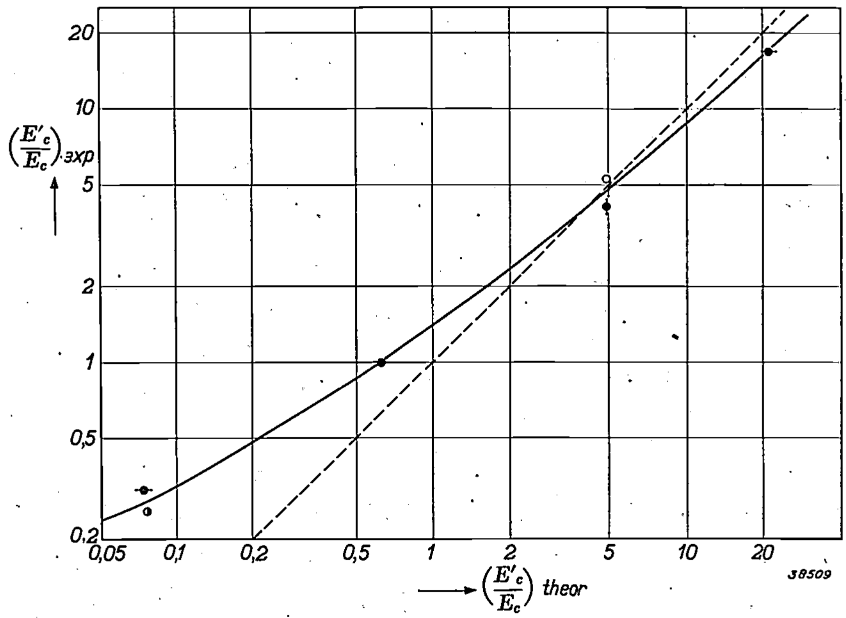


Fig. 3. The determination of the threshold values for coloured point sources of light. The ratio $(E_c':E_c)_{\text{theor.}}$ is first calculated from the threshold values of the eye illumination for coloured light and for ordinary electric light, assuming that for different colours the same number of "Nox" is required. With the help of the full drawn curve the true ratio $(E_c':E_c)_{\text{exp.}}$ of the threshold values is determined from this. The method may only be used for sources of light which emit the greatest part of their radiation in a narrow spectral region.

order to obtain from the laboratory values the practical threshold values [17]) must undoubtedly be lower than for white light. The actual ratio of the threshold values of the eye illumination will therefore be smaller than 4.7, but still always greater than unity, so that we remain on the safe side if in our calculations we treat the transmitted light as if it were white light.

Since here also the greatest part of the light is emitted in a relatively narrow spectral region, we may correct this value with the help of fig. 3, and we find $(E_c' : E_c)_{\text{exp.}} = 0.99$. Since with this colour we must expect for the correction factor N a value which does not deviate greatly from that for white light, it is also permissible here to neglect the influence of the colour of the light on the visibility.

MAGNETIC OIL FILTERS

by J. A. HARINGX.

66.067.3 : 621.318

Filters in which the oil is made to flow through a magnetic field have been constructed for the removal of particles of iron from circulating lubricating oil. In this article a description is given of such a filter which has in addition to a good efficiency in capturing iron particles many other advantages also, such as very low pressure loss in the oil circuit, no chance of the filter becoming clogged, ease in cleaning the filter when it is saturated. The action of this type of filter is studied. In ordinary use, where new particles of iron are continually entering the oil due to wear on the parts of the machine lubricated, an equilibrium is found to occur which is characterized by a certain residual concentration of iron in the oil. As the filter gradually fills up, the equilibrium is shifted to higher concentrations. If a certain limit is fixed for the residual concentration, the time can be calculated which may elapse before the filter must be cleaned. By means of a series of tests the necessary data have been obtained. The results are elaborated into a graph in which the influence of the viscosity of the oil is taken into account.

In the pressure lubrication of bearings, gears, etc. it is very important to keep the circulating oil as clean as possible. If the oil is contaminated, for instance by fine particles of iron from the wear on machine parts which rub against each other, the formation of a continuous film of oil is not only prevented, so that the lubricating power decreases, but in addition the particles cause extra wear by grinding against the parts in question. Furthermore finely divided metals, especially iron, by catalytic action, cause acidification of the oil, which is also a disadvantage for the lubricating properties.

In addition to the old textile or copper gauze filters which are used for the purification of circulating lubricating oil, magnetic filters have recently become more and more common. A magnetic filter of simple construction was described several years ago in this periodical [1]), while several possibilities of application were also discussed at that time. In the meantime a new filter of improved construction (type No. 7 715) has been developed [2]) and a series of investigations on the functioning of these oil filters has been carried out. In the following article the filter and the tests referred to are described.

The construction

In order to be able to explain the essential features of the new construction, the simplest way is perhaps to consider first the construction described previously. This is shown in *fig. 1.* In an iron housing a permanent magnet is so fastened that a magnetic circuit with a ring-shaped air gap is formed. The oil to be purified flows through this air gap. The construction is such that the stream lines of the oil and the lines along which the particles of iron are drawn to the magnet intersect each other at the smallest possible angles; by this means a relatively small lateral acceleration of the iron particles in the oil is already enough to take them out of the current and to the surface of the magnet.

Although this filter had a very satisfactory cleaning power and was able to take up large quantities of iron filings, in practice various undesired

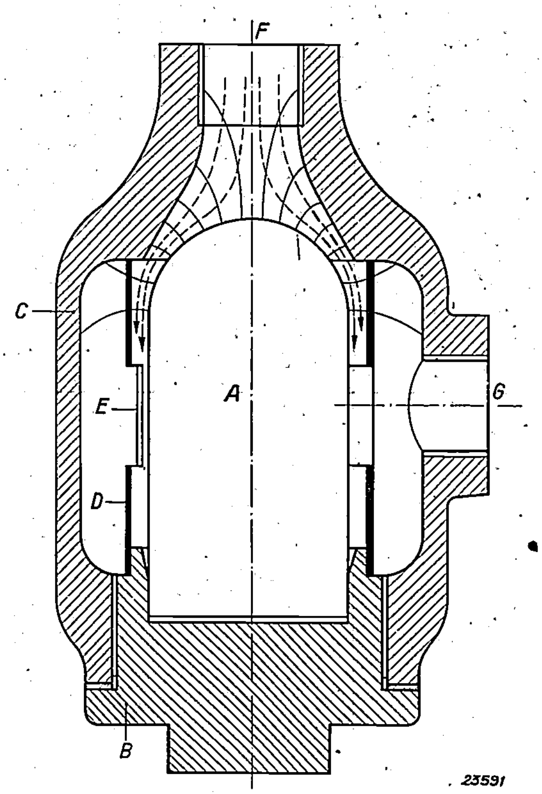
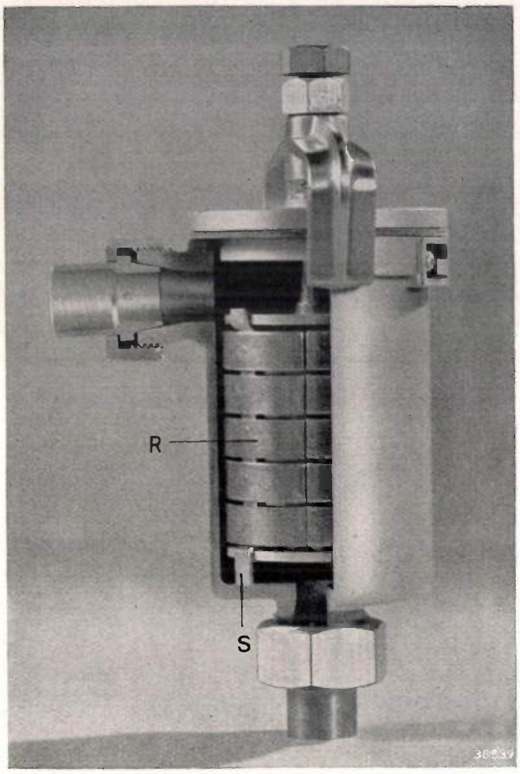


Fig. 1. Magnetic oil filter of the old type. The magnetic circuit is formed by the permanent magnet *A*, the iron cap *B* and the iron housing *C*. The oil entering at *F* flows through the air gap between *A* and *C* past the holes *E* in the copper cylinder *D* and leaves the filter at *G*. The stream lines of the oil are indicated by broken lines, the magnetic lines of force are drawn as full lines.

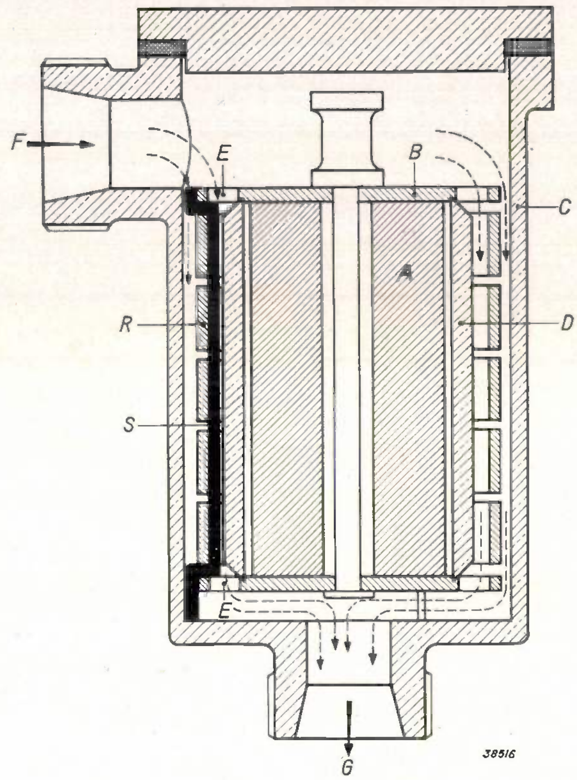
[1]) L. H. de Langen, Philips techn. Rev., 2, 295, 1937.
[2]) J. B. Aninga, Polytechn. Wbl., 33, 86, 1939.

phenomena were found to occur when the filter was approaching the limit of its capacity of retention. Due to the deposition of the iron the available opening for the flow of oil, which in connection with securing a strong magnetic field could not be made all too wide, became narrower and narrower, with the result that on the one hand a considerable pressure loss occurred in the filter, while on the other up the requirement the above-mentioned disadvantages could be avoided.

The new construction is shown in *fig. 2a* and *b*. The magnetic circuit is built up of a cylindrical magnet provided with two disc-shaped pole pieces and five soft iron rings placed around the magnet one above the other at small distances apart. In the intermediate spaces — the air gaps — strong



a



b

Fig. 2. Magnetic filter of new construction. In the cross section diagram (*b*) may be seen the permanent magnet *A* which with the pole pieces *B* and the five rings *R* forms the magnetic circuit. The oil entering at *F* flows *via* the holes *E* in the pole pieces past the inside and outside walls of the set of rings, and thus not through, but along the air gaps. The housing *C* and the cylinder *D* are of non-magnetic material. The oil leaves the filter at *G*. The rings are in two sections, as may clearly be seen in the photograph of the opened filter (*a*), and are held in place by brass strips *S* welded to them. The whole magnetic system also rests upon the ends of these strips. After unscrewing the cover of the housing the magnet system can be drawn out of it by means of the handle provided.

hand a wisp of the deposited material was sometimes torn away by the oil and passed into circulation again. Furthermore the cleaning of the filter when full was fairly difficult, since the magnet retains the deposit very firmly, a fact with which it may not be reproached since that is its function.

In the new construction the idea that the stream lines of the oil and the lines of attraction of the iron particles should intersect at only small angles has been given up. Considering the high strengths of the magnetic fields which can be obtained with the magnet steels available at present this condition is indeed no longer so important, and by giving

magnetic fields occur. The oil now, however, flows not through but along the inside and outside of the air gaps, as illustrated in *fig. 3*. It may be seen that it is actually the spread lines of force of the magnetic field which draw the iron particles floating in the oil out of their original paths and cause them to enter the air gaps. Due to the fact that the iron particles are now deposited outside the stream of oil, the tearing off of bits of the deposit has been made impossible, while at the same time no appreciable increase in loss in pressure can occur as the filter gradually fills up. In the tests to be described below the pressure loss, with a flow of 500 l/hr and a viscosity of the oil of 35 cp [3]), was found to increase during use only from 0.05 atm. in empty condition to 0.10 atm. in practically full condition of the filter. A pressure loss of this small magnitude is of no practical significance; at least,

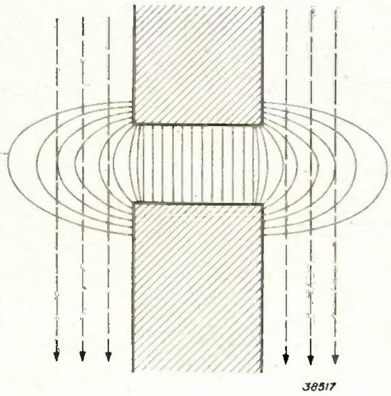


Fig. 3. Sketch of the stream lines (broken lines) and the magnetic lines of force (full lines) in the neighbourhood of one of the air gaps.

compared with the pressure losses at this rate of flow in textile filters, it may be neglected. The behaviour of the pressure loss for other values of the viscosity may be seen in *fig. 4* where curve *a* holds for the empty state and curve *b* for the practically full state of the filter.

The cleaning of the filled filter could also be made very much easier with the new construction than with the old. For this purpose each of the rings is made in two sections which are held free in position by brass strips. For cleaning, the whole magnet system is removed from the housing and the halves of the rings can easily be drawn off the magnet. Due to the fact that the rings are now no longer magnetic the deposit can be rinsed off without difficulty with some liquid or other such as petrol, trichlorethane or the like. To prevent particles of iron being deposited on the permanent magnet itself a non-magnetic cylinder is introduced between the magnet and the rings (*D* in fig. 2*b*).

The quantity of deposit which the filter can take up is proportional to the total volume of the air gaps. The larger this volume is, however, the lower the magnetic field strength which can be obtained. A suitable compromise had therefore to be found. With the chosen dimensions of the air gaps (1.5 mm long, 275 mm^2 cross section, thus a total volume of 6×415 = about 2 500 mm^3) a total of about 10 grams of iron filings can be taken up, while a magnetic field strength of 8 000 gauss is obtained.

A second compromise was necessary as to the cross section of the channel of flow for the oil. This channel must be made narrow so that the iron particles to be captured pass as close as possible along the air gaps. On the other hand less oil per hour can be dealt with when the channel is narrow, if the velocity of flow of the oil is not to be increased,

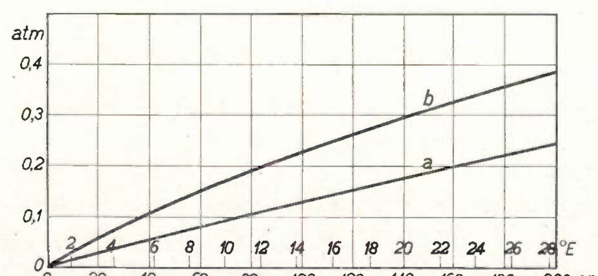


Fig. 4. Pressure loss in the magnetic filter with a rate of flow of 500 l/hr as a function of the viscosity of the oil in cp and °E, respectively. Curve *a* holds for the clean filter, *b* for the filter filled with 7 g of deposit.

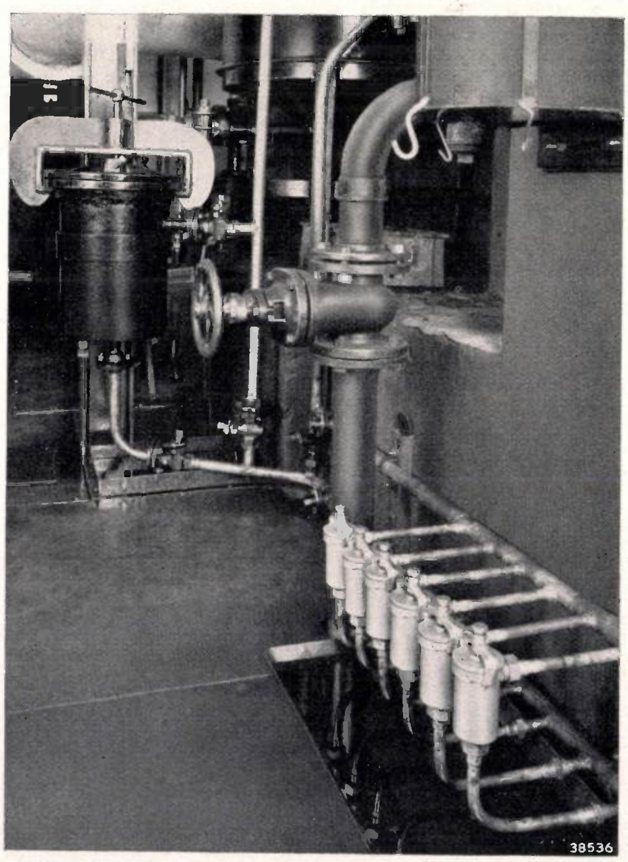


Fig. 5. Part of the lubrication system of an 800 h.p. Diesel engine. In the foreground six magnetic filters may be seen which are connected in parallel in the oil line. In the cylindrical tank on the left in the background there is a copper gauze filter which serves to remove coarse impurities from the oil.

---

[3]) A liquid has a viscosity of 1 poise (p) or 100 centipoise (cp), when a shearing stress of 1 dyne/cm^2 is necessary to maintain in the liquid a velocity gradient of 1(cm/sec)/cm. At 20°C for instance water has a viscosity of about 1 cp, rape oil one of about 100 cp. Technically the viscosity is usually given on an empirical scale (Engler degrees, °E) which is also indicated on the abscissa of fig. 4. The recalculation into cp is impossible with strict accuracy.

thereby increasing the chance that some iron particles escape being captured by the filter. The dimensions finally chosen (cross section 550 mm²) make it possible to deal with 500 litres of oil per hour. In special cases where greater flow capacities are necessary several filters may be connected in parallel in the oil line. An example of such an installation with six filters is reproduced in *fig. 5*.

Further consideration of the action of the filter

When the magnetic filter is installed in a closed system in which oil circulates which is very much contaminated by iron, the oil is seen to become clear only gradually, for instance within fifteen minutes or a half hour. This shows that all the iron particles are not caught by the filter at once as they pass it, but a certain fraction only each time: the oil must pass repeatedly through the filter before the concentration of iron falls practically to zero. If a certain quantity of new iron is continually formed in the system, the concentration is found never to fall below a certain value.

As we shall see this limiting value of the concentration is indeed very much lower than the concentration which would be present in the oil without the use of the filter; but the residual concentrations cannot in general be entirely neglected. It was therefore desirable to investigate the action of the filter more closely.

The residual concentration

Fig. 6 shows diagrammatically the situation in which the filter is used. The lubricating oil, which is kept in circulation by a pump, comes from a reservoir, passes first through the filter, then along the part of the machine to be lubricated and finally returns to the reservoir.

Suppose that the oil in the reservoir has an iron concentration of c_1. If V is the amount of oil flowing through the filter per unit of time, the filter must deal with an amount of iron particles equal to $c_1 V$ per unit of time. As a first approximation we may now assume that the same fraction of this quantity brought to the filter is always retained by it. The amount of deposit already caught by the filter, which we call G, thus grows per unit of time by the amount

$$\Delta G = \gamma\, c_1 V, \quad \ldots \ldots \quad (1)$$

where γ is a factor smaller than unity which we shall in the future designate as the retention coefficent.

If Q is the amount of new sediment formed per unit of time, Q may be larger or smaller than ΔG or equal to ΔG. When $Q > \Delta G$ more sediment is introduced into the circulating oil than is removed from it. The concentration c_1 of the iron will therefore gradually increase, at the same time, however, according to (1) the amount captured per unit of time, ΔG, also becomes larger until the addition of iron to and its removal from the oil are just balanced, *i.e.* until $Q = \Delta G$. In the same way when $Q < \Delta G$ the concentration c_1 will gradually fall until ΔG has also fallen so far that $Q = \Delta G$. It is therefore clear that in any case after some time a condition of equilibrium will be established where

$$\Delta G = Q = \gamma\, c_1 V \quad \ldots \ldots \quad (2)$$

From this we can calculate the residual concentration c_1 which may be expected. We must, however, keep in mind that the concentrations of the iron preceding and following the filter are not equal. Of the amount of iron entering the filter $c_1 V$ a part $\gamma c_1 V$ is held by it. Thus only the amount $(1-\gamma) c_1 V$ leaves the filter, *i.e.* the concentration c_2 of the iron in the oil conducted to the part of the machine to be lubricated amounts to

$$c_2 = (1-\gamma)\, c_1 \quad \ldots \ldots \quad (3)$$

From this with (2) it follows that

$$c_2 = \frac{1-\gamma}{\gamma}\,\frac{Q}{V} \quad \ldots \ldots \quad (4)$$

When the filter has functioned during a time t in the equilibrium state, with a constant addition (Q) of new sediment to the oil, it has taken up a total quantity of iron

$$G = Q \cdot t \quad \ldots \ldots \quad (5)$$

To give a numerical example: suppose that in the machine Q = 5 mg/hr of new iron filings are formed and that the retention coefficient has the apparently very low value of $\gamma = 0.01$. If we substitute for V the above-mentioned value of 500 l/hr, c_2 becomes only 1 mg/l.

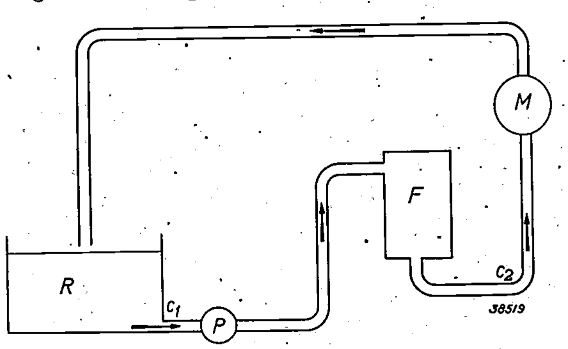


Fig. 6. Diagram of a lubricating system with oil circulation. R reservoir, P oil pump, F magnetic filter, M part of the machine to be lubricated. The oil has an iron concentration of c_1 preceding the filter and iron concentration of c_2 after passing it.

With a retaining capacity of $G_{max} = 7$ grams, according to (5) the filter must be cleaned after 1 400 hours of use. Without the filter after 1 400 hours, when a total of 50 litres of oil take part in the circulation there would be an iron concentration of 140 mg/l [4]).

Further consideration of the retention coefficient

How large is the retention coefficient γ of the filter described? It is impossible to give an immediate answer to this question since the above assumption of a constant value of γ is by no means justified. The retention coefficient depends upon a number of factors, the chief of which are: the degree G to which the filter is filled, the iron concentration c_1 in the oil, the viscosity of the oil, the rate of flow and the size of the iron particles. The last three factors remain approximately unchanged during use of the filter, or at least they may be kept constant in the experiments to be described in the following. This is, however, not true of the first two, so that we must extend our consideration of this point somewhat. We shall first discuss the influence of the degree G to which the filter is filled.

When the filter is entirely filled with the deposit γ must in any case fall to zero. This transition to zero does not take place abruptly, which is understandable when it is kept in mind that the magnetic field itself is influenced by the iron particles taken up. The field strength will become greater due to the decrease in the magnetic resistance of the air gaps. At the same time, however, the spreading, which is just what we must have, will become less. This last effect is found to dominate and for the behaviour of γ with G a curve of the form sketched in *fig. 7* is obtained.

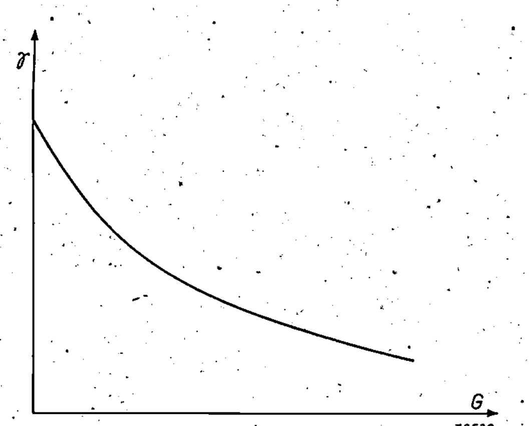


Fig. 7. Approximate form of the curve showing the variation of the retention coefficient γ with the degree of filling G of the filter.

[4]) In general before such high concentrations are reached the oil will have to be renewed.

The result of this variation of γ with G is that, strictly speaking, we may no longer speak of a definite equilibrium condition: during use, as the filter slowly fills up, γ will decrease slowly and the iron concentration in the oil will gradually increase. Nevertheless upon closer consideration it is found that when γ does not vary too much with G, a condition which is always satisfied in use, the successive states of the system may still be conceived of as a gradually shifting equilibrium. This means that we may continue to use equation (4) for the calculation of c_2, with a different value of γ at every moment corresponding to the degree of filling G of the filter at that moment.

If the requirement is made that the residual concentration c_2 of the iron may not exceed a certain value, it includes the requirement that the filter must be cleaned, not when it is full, but sooner, namely when the retention coefficient has fallen to that value of γ which according to (4) corresponds to the permissible value of c_2. From a graph like that of fig. 7 one may then read off at what degree of filling G the filter should be cleaned, and from equation (5) the time can be calculated which is necessary to reach this degree. Equation (5) is no longer exactly valid, since due to the slow shift of the equilibrium a certain amount of the newly formed sediment is used in increasing the iron concentration of the oil. Nevertheless, considering the low concentration and the small quantities of oil which are usually used for the circulation, the deviation from equation (5) is only slight.

The second factor which varies during use and which affects the retention coefficient is the concentration c_1 itself. Its effect is easily understood qualitatively. The diagram of the lines of force sketched in fig. 3 will be somewhat altered when the magnetic resistance is decreased on both sides of the air gap proper by the presence of a large number of iron particles. The "spreading" increases as it were with the iron concentration of the oil, and at the same time the attractive effect of the field also increases. This effect is found to be quite considerable in magnitude.

The fact that γ depends upon c_1 can be taken into account in quite the same way as above. For every state characterized by a definite value of G and of c_1 there is a certain value of the retention coefficient γ. If we represent this relation graphically by plotting γ as a function of G for a series of values of c_1 (*fig. 8*), we can again read off the degree of filling G of the filter at which it must be cleaned, when a given maximum residual concentration c_2 is prescribed. γ can be calculated from c_2 according

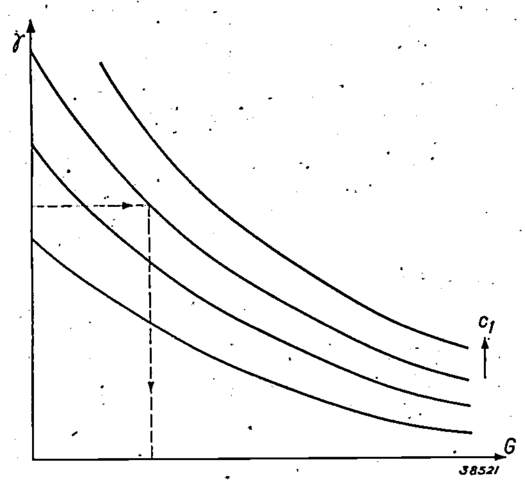


Fig. 8. Approximate form of the family of curves which indicates the retention coefficient γ as a function of the degree of filling G with as parameter the iron concentration c_1 at the inlet to the filter. If a given residual concentration c_2 is permissible at the outlet of the filter, with the aid of equations (4) and (3) γ and c_1 can be calculated, and the permissible degree of filling can then be read off in this graph.

to (4) and c_1 according to (3). We shall see below how this graph can be drawn somewhat more simply; the principle, however, remains the same.

Experimental determination of the retention coefficient

It is now a question of studying more closely the relation γ (G; c_1) sketched in fig. 8 for our filter, which can only be done experimentally. The necessary experiments are preferably related as much as possible to practical conditions. One of the conditions should therefore be a constant and regularly distributed addition Q of iron particles to the oil, and in order to include the whole range of degrees of filling the experiment for each value of Q (with the corresponding series of concentrations c_1 which are traversed) should be continued for a long time, as is apparent from the numerical example described above. In order to avoid this we have arranged the experiments somewhat differently. A certain quantity (H) of iron filings was added once only to the circulating oil (total amount v). In contrast to the usual case in which the iron concentration increases very slowly as the filter becomes full, we here have a fairly rapid decrease of the concentration, and the experiment is concluded in a few hours. The variation of the concentration c_1 with the time is determined by measuring c_1 chemically at specified intervals. The retention coefficient γ is then found as follows. At each moment the following is true:

$$G = H - vc_1. \quad \quad (6)$$

The amount of iron G retained in the filter thus grows per unit of time by

$$\frac{\mathrm{d}G}{\mathrm{d}t} = -v\,\frac{\mathrm{d}c_1}{\mathrm{d}t}.$$

On the other hand, according to the definition of the retention coefficient (equation (1)) this growth was given by $\gamma c_1 V$. Therefore

$$\gamma\, c_1 V = -v\,\frac{\mathrm{d}c_1}{\mathrm{d}t},$$

$$\gamma = -\frac{v}{V}\,\frac{\mathrm{d}\ln c_1}{\mathrm{d}t}. \quad \quad (7)$$

Thus if we plot ln c_1 against t, γ follows from the slope of the curve obtained.

In all experiments the maximum rate of flow of $V = 500$ l/hr was used, while the viscosity of the oil, which varies very much with the temperature, was kept constant by placing the reservoir in a thermostat. The experiments were carried out with a very fine carbonyl iron powder which satisfactorily resembles the sediment most commonly formed in practice as far as shape and size of the particles (about 2 microns) are concerned. After the addition of the powder (G = 1 to 10 g in v = 5 litres of oil), the oil was first pumped through the system for several hours with no magnet in the filter, in order to distribute the iron evenly throughout the oil. Special care had to be taken that all the oil took part in the circulation, *i.e.* that there was no direct

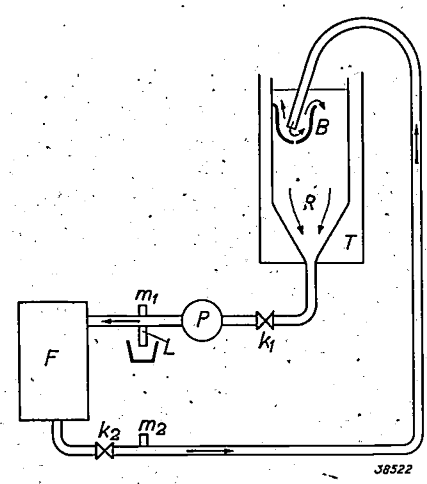


Fig. 9. Arrangement for the experiments for the determination of the retention coefficient. R reservoir, T thermostat, P toothwheel pump, F magnetic filter, k_1, k_2 taps. In order to avoid a direct current in the reservoir between inlet and outlet, which would hinder the uniform distribution of the iron throughout the oil, the inlet ends in a vessel B over the edge of which the oil flows in all directions. There is a small hole in the bottom of the vessel to prevent any iron which has settled from being removed from the circulation. At specified intervals a small quantity of oil is tapped off through the line L and its iron concentration is determined. At m_1, m_2 a differential manometer was connected which indicated the pressure loss in the filter.

current in the reservoir from inlet to outlet. *Fig. 9* shows how this was accomplished.

The apparently so simple experiments still contained many stumbling blocks. While the circulation pump is getting under way, for example, there is great danger of the formation of small air bubles in the oil, which later have no opportunity of escaping from the rapidly flowing oil and which upon passing through the filter may affect the deposition of iron particles considerably by their surface tension. By starting the pump as gradually as possible the formation of air bubbles is almost entirely avoided. Furthermore a small amount of iron powder is found to cling to the walls of the circulation system at various points, which results in the fact that the iron concentration is slightly lower than corresponding to equation (6). This can be taken into account by using a corrected value for v as well as for H. Further, the concentration c_1 in equation (6) is actually an average value over the total quantity of oil, while the action of the filter is determined by the concentration at the inlet to the filter. A correction must also be applied for this by correlating the slope (γ) found at point t in the $\ln c_1$—t curve with the average value of c_1 over an adjacent time interval τ; τ is here the time necessary for the total amount of oil to be pumped once around the system, in our case about 36 sec. Nevertheless, after all these corrections, the results still show considerable divergence, caused by the limited accuracy of the measurement of the concentration. This measurement is made by tapping from the oil line in fig. 9 a sample of 25 cc of oil, "igniting" the oil and determining the iron content of the residue.

Results of measurements

In the manner described three series of measurements were carried out in the first instance, in which $1^3/_4$, 5 and 10 g of iron powder, respectively, were added to the oil. In *fig. 10* the variation of the measured concentration c_1 is given as function of the time elapsed after the insertion of the magnet for these three cases (*1*, *2*, *3*). If we now consider for instance curve *2*, for every point c_1, t of the curve we can determine the values of γ, c_2 and G corresponding to that value of c_1 according to equations (7), (3) and (6). In *table I* the result of this determination is given for a series of values of c_1 for curve *2*.

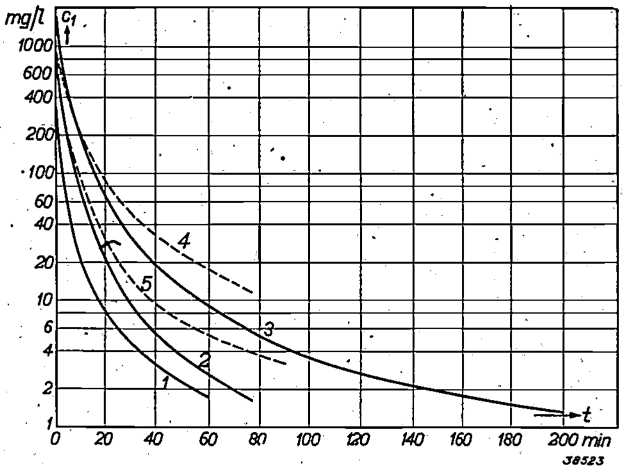


Fig. 10. Variation of the iron concentration c_1 (in mg/l) with the time t (in min) in different experiments. At the beginning of each experiment quantities of iron were added to the oil: $H = 1.75$ g (curve *1*), 5 g (curve *2*) and 10 g (curve *3*). $\ln c_1$ is plotted against t in order to find the retention coefficient γ by graphical differentiation according to equation (7). The broken line curves *4* and *5* with $H = 5$ and 4 g. respectively, were recorded with a higher viscosity of the oil, namely with $\eta = 85$ cp, while in curves *1*, *2* and *3* the viscosity was 35 cp.

According to the discussion above, for every value of the parameter c_1 we would have to plot the retention coefficient γ as a function of G. Then on the graph obtained, beginning with γ and c_1, the permissible value of G could be read off. In the practical application, however, we must assume that the values of Q (the quantity of new sediment continuously formed in the part of the machine being lubricated) and of c_2 (the permissible residual concentration) are given, and that γ and c_1 must then be calculated with the help of (4) and (3). It

Table I.

Elaboration of the series of measurements *2* and *5* (as examples). The rows with large numbers are derived directly from the curves *2* and *5*, respectively in fig. 10, while the rows with small numbers are found by interpolation.

Experiment 2					Experiment 5					
c_1 mg/l	γ	c_2 mg/l	G mg	$\frac{\gamma\, c_2 V}{1-\gamma}$ g	c_1 mg/l	γ	c_2 mg/l	G mg	$\frac{\gamma\, c_2 V}{1-\gamma}$ g	$\frac{\gamma\, c_2 V}{1-\gamma}\,\frac{\eta'}{\eta}$ g
786	0.260	580	0		753	0.152	638	0		
	0.237	500	830	77.6		0.141	500	1070	41.0	100
400	0.196	321	2410		400	0.126	350	1868		
	0.110	100	3853	6.16		0.081	100	3180	4.40	10.8
50	0.086	46	4187		50	0.0585	47	3450		
	0.068	20	4343	0.730		0.0415	20	3590	0.434	1.065
20	0.068	18.6	4350		20	0.0415	19.2	3594		
10	0.052	10	4402	0.274		0.0263	10	3640	0.135	0.331
5	0.036	5	4431	0.0931	10	0.0260	9.75	3641		
	0.0250	3	4443	0.0384		0.0123	5	3665	0.0311	0.0763
2	0.0172	2	4449	0.0174	3	0.0056	3	3675	0.0084	0.0206

is therefore simpler to make the graph in such a way that the given quantities, in addition to the required value of G, occur in it directly. In order to do this we choose c_2 as parameter with the values of $1^1/_2$, 2, 3, 5, 10 ... mg/l, for example. In table I several of these values are filled in by intrapolation (by means of auxiliary graphs) and in each row the quantity

$$\frac{\gamma c_2 V}{1-\gamma} = Q \quad \ldots \ldots \quad (8)$$

is calculated and indicated in the fifth column. (For the rate of flow V we always use the maximum value of 500 l/hr, with which the experiments were also performed). If we now plot this quantity as a function of G with c_2 as a parameter, the required value of G can be read off directly from the given values of Q and c_2.

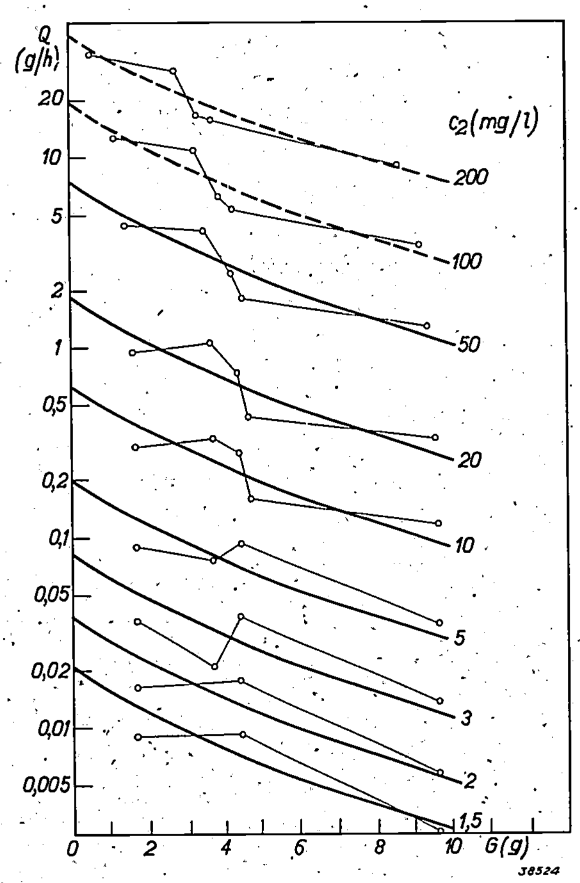


Fig. 11. The quantity $\gamma c_2 V/(1-\gamma)$, which is equal to Q in the equilibrium state of normal functioning, is here plotted as a function of G with c_2 as a parameter. For every curve c_2 there are three measured points available from the series of measurements *1*, *2*, *3*, while two series of measurements *4*, *5* carried out with a different viscosity of the oil provided two more points for every curve. (For the broken line curves the quantity $\gamma c_2 V/(1-\gamma)$ here plotted may not be set equal to Q, since at such high concentrations the change in the quantity of iron present in the oil may no longer be neglected. The curves are, however, of scarcely any importance, and are here given only for the sake of fixing more completely the shape of the curves of the whole family).

For each curve for one value of c_2, each of the three series of measurements *1*, *2*, *3* gives one point (a pair of values of Q and G), so that we would have to draw each curve with the aid of only three measured points. This would be quite difficult considering the fairly great divergences. Two other series of measurements (see below) performed by a slightly different method, however, provided two more points for every curve, so that the shape of the curves was somewhat better determined. The family of curves obtained is reproduced in *fig. 11.*

The influence of the viscosity

While in the above in the theoretical as well as the experimental considerations we have only investigated the influence of the degree of filling G and the iron concentration c_1 on the retention coefficient, we shall in conclusion also study the influence of the viscosity of the oil. The fact that the viscosity will influence the action of the magnetic filter is easily understood: the more viscous the liquid, the greater the force necessary to remove a particle from it. At a given rate of flow the magnetic field will capture the iron particles from a thick oil less well than from a thin oil.

As in our experiments, we may also assume that for normal use in a given installation, in which the working temperature and type of oil chosen are fixed, the oil will have a constant viscosity. A graph like fig. 11 can therefore be used, with a separate curve for each viscosity. Actually, however, there is a still simpler method.

The three series of measurements *1*, *2*, *3* in the experiments with the arrangement described were obtained with a single definite viscosity of the oil, namely η = approx. 35 cp. Two additional experiments (*4* and *5*) were performed with a viscosity of η' = approx. 85 cp. The variation of the measured concentration with the time for these latter cases, where H = 5 and 4 g, respectively, is given as a broken line in fig. 10. It may immediately be seen that the retaining action of the filter is smaller since the iron concentration decreases more slowly. If, however, the retention coefficient γ as well as the values of c_2, G and $\gamma c_2 V/(1-\gamma) = Q'$ (see table I where this has been done for curve *5*), are again calculated from the curves, and if again as above the points of a family of curves as in fig. 11 are determined from these values, these points are found to fit very well in the old family of curves when the values of their ordinates Q' are multiplied by the ratio of the viscosities $\eta'/\eta = 85/35$. The agreement — as far as may be expected with the fairly large divergences already mentioned — is so good that we used the two new points obtained in this way for every c_2 curve to draw a smooth curve through the points.

The result is therefore that the same family of curves is obtained if for experiments with the viscosity η' the quantity $\eta' Q'$ is plotted and for experiments with the viscosity η the quantity $Q\eta$. We have therefore redrawn the graph of fig. 11 with this ordinate, see *fig. 12*, so that it may be used for any desired viscosity. At the same time a number of interpolated curves are given in this figure.

A numerical example will serve to illustrate the use of the graph. A large gear box in which about 25 mg of sediment is formed per hour (this apparently very high figure actually does occur in practice) is lubricated with 1 000 litres of oil per hour. Since the maximum capacity of flow of the magnetic filter here described is 500 l/hr, two filters in parallel must be installated in the oil circuit. At the working temperature the oil has a viscosity of 50 cp. The problem is to determine when the filters must be cleaned if the iron concentration may not exceed 2 mg/l (in general a considerably greater concentration will be permissible in a gear box).

Per filter the amount of new sediment introduced is $Q = 12.5$ mg/hr. If in fig. 12 we intersect the horizontal line corresponding to $0.0125 \times 50 = 0.625$ by the curve corresponding to $c_2 = 2$ mg/l, then at the abscissa of the point of intersection it may be read off that each filter may become filled to a weight of $G = 3$ grams of iron particles. The time necessary for this is $3/0.0125 = 240$ working hours. After that time the filters must be cleaned. (If one waits for instance 600 instead of 240 hours G becomes 0.5 g and according to fig. 12 the residual concentration has still only increased to 3 mg/l).

If it were desired in the above case to keep the iron concentration below 2 mg/l without the use of filters, then per working hour $25/2 = 12.5$ litres of fresh oil would have to be added. This illustrates very clearly the advantage of the use of such magnetic filters.

In conclusion a few words may be said about the permissible value of the iron concentration c_2 (residual concentration). This quantity, like the amount of sediment Q, will vary very much according to the nature of the part of the machine lubricated; with accurately finished axles and bearings c_2 must be small, for instance not more than several mg/l. If at the same time a high value of Q and/or η occurs it might be that the corresponding horizontal line in fig. 12 lies in its entirety higher than curve c_2. This would mean that the retention coefficient of the magnetic filter is from the very first unable to keep the residual concentration below the desired value. In such a case recourse may be had to two or more magnetic filters connected in series in the oil line — solution which is possible because of the very slight pressure loss of these filters already mentioned. Experience gained until now, however, indicates that for by far the majority of cases the retention coefficient of one filter is more than adequate.

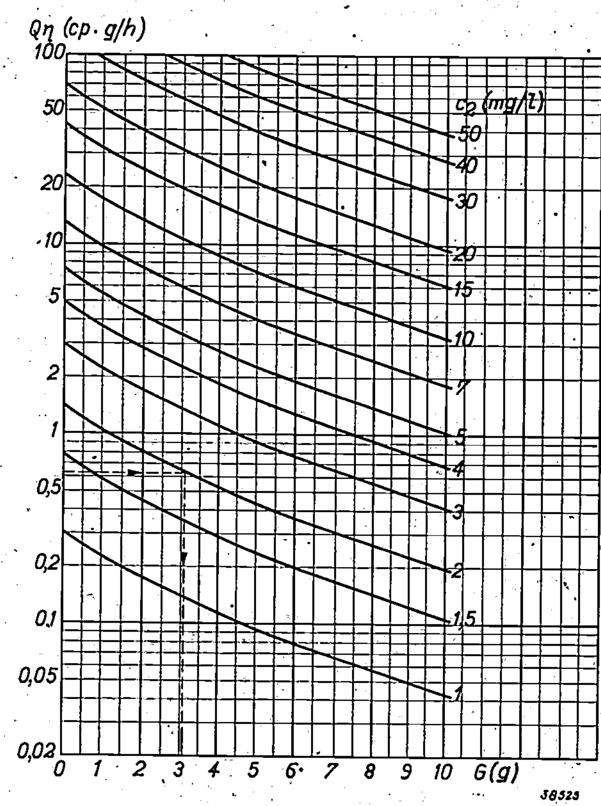


Fig. 12. Graph for use in the practical application of the magnetic filter described. With a given viscosity η of the oil (in cp), a given quantity of new sediment Q (in g/hr) and a given permissible residual concentration c_2 of the iron (in mg/l), it can be read off on this graph to what degree G (in grams of iron) the filter may become filled before it must be cleaned. G/Q is then the corresponding number of working hours. The graph is valid for a rate of flow of 500 l/hr and for iron particles 2 microns in size.

THE NOISE IN RECEIVING SETS AT VERY HIGH FREQUENCIES

by M. J. O. STRUTT and A. van der ZIEL.

621.396.645.3 : 621.396.822

As a result of the spontaneous current and voltage fluctuations which occur in the input circuit of a receiving set, very weak aerial signals cannot be amplified without interference. The relation between the fluctuation currents and the signal currents in the anode circuit of the first amplifier valve may be taken as a measure of the interference. In this article the factors are investigated which determine the interference with the reception at a given energy of the signal. It is found that at relatively low frequencies (broadcasting frequencies) the fluctuations need not interfere with the reception. At very high frequencies for the case of a triode as first amplifier valve it is found that the interference is determined by the ratio between the noise resistance and the electron-input resistance of the valve. The damping due to the self-induction of the cathode connections has no influence on the ratio between fluctuation current and signal current. Finally the calculations are carried out for pentodes as well. It is shown that the so-called distribution fluctuations may be compensated by introducing a suitable self-induction into the screen grid connections, As to the other fluctuations, the behaviour of a pentode does not differ fundamentally from that of a triode.

When the sensitivity of a receiving set is raised higher and higher, for example by connecting a larger and larger number of stages behind one another, weaker and weaker signals can be received with the desired output amplitude. At the same time, however, it will be found that the reception experiences a larger and larger degree of interference by the familiar noise.

The noise in receiving sets is of fundamental significance in their practical use. *The lower limit of the intensity which an aerial signal must have in order to give satisfactory reception is not determined by the fact that at still lower input voltages the acoustic output signal falls below the threshold value, but by the fact that at still lower input voltage the acoustic output signal is drowned out by the noise.*

The causes of noise in amplifiers have already been discussed several times in this periodical [1]). As sources of noise were pointed out spontaneous current and voltage fluctuations which may be partially ascribed to thermal causes (Brownian movement) and which are partially connected with the discontinuous corpuscular nature of the electric charge. The weaker the signal which must be dealt with by an amplifier stage, the relatively greater the interference caused by these spontaneous fluctuations. If the ratio between the fluctuation amplitude and the signal amplitude is considered as a measure of the disturbing effect, and if for example the first amplifier stage has a tenfold amplification, the spontaneous fluctuations in the second stage have only 10 per cent of the disturbing effect of those in the first stage. If the fluctuation amplitudes are the same in both stages, then due to the fluctuations in the second stage the total fluctuation amplitude increases by a factor $\sqrt{1^2 + 0.1^2}$, and therefore becomes only 0.5 per cent greater. The fluctuations in the third and following stages are practically entirely without significance. It is therefore in general the noise in the first and sometimes the second stage which determines the sensitivity of a receiver.

A complete calculation of the spontaneous fluctuations in the first two stages was given some years ago in this periodical [2]). This treatment is valid only for ordinary and short broadcasting waves. With waves shorter than for instance 10 m, however, these considerations are no longer valid, since at the corresponding frequences the capacitative and inductive characteristics of the electrode system and of its wiring begin to exert an appreciable influence on the action of the electronic valve. At about the same frequencies the transit time of the electrons between cathode and control grid also begins to affect the action of the valve and especially the noise phenomena [3]). Both of these effects must be taken into account in the calculation of the sensitivity of an amplifier stage for waves shorter than 10 m. A few extentions of the calculation of the fluctuation phenomena in a receiving set resulting from these effects form the subject of this article.

---

1) M. Ziegler, Philips techn. Rev. 2, 136, 329, 1937. C. J. Bakker, Philips techn. Rev., 6, 129, 1941. See also M. J. O. Strutt Spontane spannings- en stroomfluctuaties (ruischen) in electronenbuizen en aangesloten ketens. (Spontaneous voltage and current fluctuations (noise) in electronic valves and circuits in connection with them), T. Ned. Radio Genootsch. 9, 1, 1941.

2) M. Ziegler, Philips techn. Rev., 3, 189, 1938.

3) This was pointed out by the authors in the discussion of the modern high-frequency amplifier valves EFF 50 and EF 51, Philips techn. Rev. 5, 172 and 357, 1940.

Sources of fluctuation in an amplifier circuit

Fig. 1 shows the principle of a high-frequency amplifier stage in which, for the sake of simplicity, it is assumed that the amplifier valve is a triode. A is an aerial with an internal resistance (radiation resistance) R_{ant}. The voltages excited in this aerial by the incoming signal are applied, *via* the ideal transformer T, to an oscillation circuit which is equivalent in tuning to a resistance R_{LC}. A voltage occurs on this resistance which acts between grid and cathode of the triode, and causes in this way an anode alternating current.

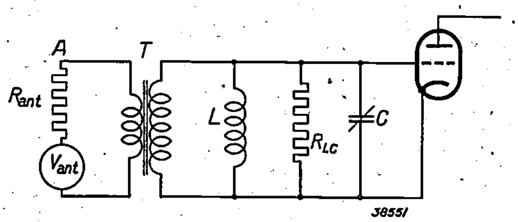


Fig. 1. Principle of the input connections of a receiver. A aerial with internal resistance (radiation resistance) R_{ant}, T ideal transformer with transformation ratio t, L-C oscillation circuit with resonance resistance R_{LC}.

In addition to the signal voltage V_{ant} indicated in the diagram, interference voltages and currents occur at various spots in the circuit due to the fluctuation phenomena mentioned above. Together these lead to certain current fluctuations in the anode circuit of the amplifier valve. *The ratio between the fluctuation current and the signal current in the anode circuit will in the following be considered as a measure of the quality of the reception. The suitability of a valve or of an amplifier connection for the amplification of small signals may be considered greater, the smaller the aerial voltage with which a given quality of reception can be obtained.*

In order to ascertain what factors determine the quality of reception with a given aerial voltage V_{ant}, we shall consider more closely the different sources of voltage and current fluctuations. Four sources of interference may be distinguished in the connections:

1) a disturbance of the aerial signal by other interfering waves in space (v_{ant}),
2) an interference voltage v_C in series with the resonance resistance of the oscillation circuit,
3) an interference current i_k in the anode circuit due to the spontaneous fluctuations in the number of electrons and in the velocity at which they are emitted from the hot cathode of the valve,
4) a control grid current i_g correlated with the above which results from the fact that the electrons which are in the space between the electrodes cause the appearance of an induction charge on the control grid which alternates in the rhythm of the anode current fluctuations.

We shall not go into details about these sources, but shall confine ourselves to giving the formulae with which the fluctuation voltages and currents of fig. 1 can be calculated. For a detailed derivation of these formulae we refer to the article of C. J. Bakker cited in footnote [1]).

1) Interfering waves in space

The interference voltage on the aerial which occurs within a frequency region Δf due to interfering waves in space is given by:

$$\overline{v_{ant}^2} = a \cdot 4\, kT\, R_{ant}\, \Delta f, \quad . \quad . \quad . \quad . \quad (1)$$

where $k = 1.38 \times 10^{-23}$ W sec/degree is the Boltzmann constant, while T is room temperature in °K and a is a numerical factor lying between 1 and 10^4).

2) Interference voltage in series with the oscillation circuit.

For the fluctuation voltage in series with the resonance resistance of the oscillation circuit the following relation is valid in a narrow frequency region close to the resonance frequency:

$$\overline{v_C^2} = 4\, kT\, R_{LC}\, \Delta f, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If Δf is not a very narrow frequency region, R_{LC} must be defined more precisely, namely as the average value of the real part of the circuit impedance for the frequencies in question. In practice, however, this will not usually involve a very great difference.

3) Spontaneous current fluctuations in the anode circuit of the triode

The spontaneous current fluctuations in the anode circuit of a triode may be described by the formula:

$$\overline{i_k^2} = F^2 \cdot 2\, e\, I_k\, \Delta f, \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

where $e = 1.6 \times 10^{-19}$ coulomb represents the charge on an electron. I_k is the current which leaves the cathode and F a factor, the so-called noise factor.

For negative control voltages (starting current region) and for control voltages which have such

[4]) The exact value of a depends upon the directional characteristic of the aerial. The radiation does not come equally from all parts of the sky, but mainly from the direction of the milky way. If an aerial with a sharp directional effect is used and directed toward the milky way, values of a in the neighbourhood of 30 are found. See in this connection the article referred to in footnote [1]) by C. J. Bakker in which the literature is given.

a large positive value that saturation occurs, $F = 1$. In the practical working range of the amplifier valve in which the anode current is limited by space charge, F is much smaller than unity. The behaviour of F in this region is such that the fluctuations can be represented in approximation by the relation:

$$\overline{i_k^2} = 4\,k\,T_k\,S\,\Delta f \quad . \; . \; . \; . \; . \quad (3b)$$

where T_k is the cathode temperature and S the slope of the triode [5]).

Another comprehensive method of representing the current fluctuations i_k is the introduction of the concept of noise resistance. The noise resistance R_r of a valve is defined by the relation:

$$\overline{i_k^2} = 4\,kT\,R_r\,S^2\,\Delta f \quad . \; . \; . \; . \; . \quad (3c)$$

where T is room temperature (290 °K). R_r is the resistance which when connected between cathode and control grid of an electronic valve causes by its thermal voltage fluctuations at room temperature just as large current fluctuations as actually occur spontaneously. From equations (3b) and (3c) R_r can easily be calculated. For a cathode tempe-rature of 770° C (1 040 °K) we find:

$$R_r \approx 4/S \quad . \; . \; . \; . \; . \; . \; . \quad (4)$$

4) Spontaneous current fluctuations in the control grid circuit of the triode

The current fluctuations in the control grid circuit are induced by alternations of the electric charge which is situated in the space between the electrodes. Their behaviour is the same as that of the anode current fluctuations i_k, but there is a phase shift of 90° with respect to the latter. The intensity of the current fluctuations may be expressed in the form of the input damping $1/R_e$ which occurs due to the space charge between cathode and control grid. If the space charge between control grid and anode may be neglected, the following relation holds between the current fluctuations and the electron unput damping:

$$\overline{i_g^2} = 1{,}43\,\frac{4\,k\,T_k}{R_e}\,\Delta f \quad . \; . \; . \; . \; . \quad (5)$$

If the space charge between control grid and anode is important, instead of 1.43 we have a slightly smaller factor, 1 for example.

The mean squares of the sources of fluctuation v_{ant}, v_C, i_k and i_g are determined by the equations (1, 2, 3c and 5). Together with the signal voltage V_{ant} these data are sufficient to furnish a complete picture of the behaviour of the circuit given in fig. 1.

The ratio between fluctuation current and signal current in the anode circuit of the first amplifier valve

After having discussed in the foregoing the four sources of fluctuation of a high-frequency amplifier stage, we shall now study the current fluctuations which are caused in the anode circuit of an amplifier valve in the connections of fig. 1. Since with the help of the same diagram we can also calculate the amplitude of the signal current in the anode circuit, the ratio between fluctuation amplitude and signal amplitude also follows, and in the foregoing we have considered this ratio as a measure of the quality of the reception.

In order to simplify the diagram we omit the ideal transformer T and replace the aerial resistance as well as the aerial voltages (fluctuation voltages and signal voltage) by their transformed values. If the transformation ratio of the transformer (number of secondary to number of primary windings) is equal to t, the voltages must be multiplied by a factor t and the aerial resistance by a factor t^2. In *fig. 2* the diagram so obtained is shown; the electron input resistance R_e has also been taken into account. The fluctuation voltages v_{ant} and c_C and the fluctuation current i_g cause definite variations in the control grid voltage, which when multiplied by S give the required fluctuations of the anode current. To this must still be added the spontaneous anode current fluctuations i_k, *i.e.* the accidental variations of the anode current which occur due to fluctuations in the mechanism of thermal emission at a constant grid voltage.

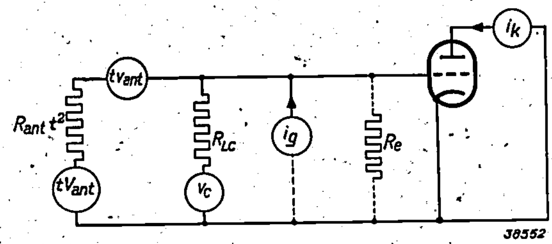


Fig. 2. Equivalent diagram of the input given in fig. 1. The transformer with the aerial connected to it is replaced by an aerial with transformed voltages and transformed radiation resistance. In addition to the aerial voltage V_{ant}, there are also the fluctuation voltages v_{ant} and v_C and the fluctuation currents i_g and i_k.

[5]) In the article referred to in footnote [1]) by C. J. Bakker, the following relation was derived for a diode:

$$\overline{i_k^2} = 0{,}64 \cdot 4\,k\,T_k\,S\,\Delta f.$$

At the same time it was noted that for a triode a similar expression is valid, except that for S a value must be taken which is 1.5 to 2 times as large as the measured slope. With the normal amplifier valves this factor is found to be about 1.6, so that the factor 0.64 is thereby increased to about 1.

In the performance of the calculations it will be found useful to make repeated use of the following theorem (see *fig. 3*). *A source of voltage V with an internal resistance R in series with it can be replaced*

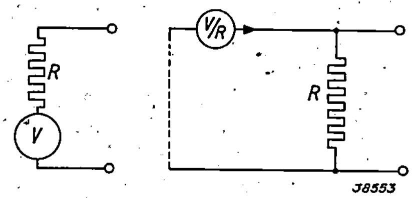


Fig. 3. A source of voltage with an EMF V and an internal resistance R corresponds in action to a source of current furnishing a current V/R to which a resistance R is connected in parallel. The internal resistance of the source of current (dotted line) must be imagined to be infinitely great.

by a source of current of the strength $I = V/R$ in parallel with a resistance R. When this theorem is used it is clear that for example the aerial voltage in fig. 2 can be replaced by a source of current of strength $V_{ant}/(R_{ant}t)$, whose current flows through the resistances $R_{ant}t^2$, R_{LC} and R_e connected in parallel. If we indicate the resistance value of this system of three resistances connected in parallel by the symbol $[R_{ant}t^2|R_{LC}|R_e]$, we find for the control grid A.C. voltage excited by the aerial signal the following:

$$V_g = \{V_{ant}/(R_{ant}\,t)\}\,[R_{ant}\,t^2|R_{LC}|R_e] \quad . \quad . \quad (6)$$

In a similar way it may be calculated that the voltage sources v_{ant} and v_C and the current source i_g give rise to the following fluctuation voltages on the control grid:

aerial fluctuations:

$$\delta_1 V_g = \{v_{ant}/(R_{ant}\,t)\}\,[R_{ant}\,t^2|R_{LC}|R_e]\,; \quad (7a)$$

circuit fluctuations:

$$\delta_2 V_g = (v_C/R_{LC})[R_{ant}\,t^2|R_{LC}|R_e]\,; \quad . \quad . \quad . \quad (7b)$$

grid current fluctuations:

$$\delta_3 V_g = i_g\,[R_{ant}\,t^2|R_{LC}|R_e] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7c)$$

The total grid voltage fluctuation is equal to the sum of the three separate contributions. If this sum is multiplied by S, the anode current fluctuation thereby caused is obtained. This is thus given by

$$\delta I_a = \delta_1 I_a + \delta_2 I_a + \delta_3 I_a =$$
$$= S\left(\frac{v_{ant}}{R_{ant}\,t} + \frac{v_C}{R_{LC}} + i_g\right)[R_{ant}\,t^2|R_{LC}|R_e].$$

If we add to these fluctuations of the anode current the fluctuations

$$\delta_4 I_a = i_k, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7d)$$

we obtain the total anode current fluctuations. Finally we may divide the fluctuation current thus obtained by the signal current I_a which is obtained by multiplying the grid voltage V_g (equation 6) by S. In this way we find for the required ratio between fluctuation current and signal current:

$$\frac{\delta I_a}{I_a} = \frac{v_{ant}}{V_{ant}} +$$
$$+ \frac{R_{ant}\,t}{V_{ant}}\left\{\frac{v_C}{R_{LC}} + i_g + \frac{i_k}{S\,[R_{ant}\,t^2|R_{LC}|R_e]}\right\} \quad . \quad (8)$$

The influence of the self-induction of the cathode connections at very high frequencies

At very high frequencies the self-inductions and mutual inductions of the connections of the valve and the wires used for the connections may have an important effect on the action of an amplifier stage, so that there may be some doubt whether fig. 2 represents a usable equivalent circuit diagram of the amplifier stage. In particular the self-induction of the cathode connections leads to the fact that the admittance of the valve between cathode and control grid is not given only by the electron input damping $1/R_e$, but is several times larger with the normal dimensions of high frequency amplifier valves. If the total damping is written in the form

$$\frac{1}{R_i} = \frac{1}{R_e} + \frac{1}{R_{L_k}}\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

then for the share of the cathode connections in the damping the following may be calculated [6]):

$$\frac{1}{R_{L_k}} = \omega^2\,L_k\,C_{kg}\,S, \quad . \quad . \quad . \quad . \quad (10)$$

where L_k is the self-induction of the cathode connections and C_{kg} is the capacity between cathode and control grid.

[6]) See Philips techn. Rev., 3, 103, 1938. The derivation is briefly as follows. If V is the A.C. voltage between the control grid and the end of the cathode pin, and if I_a is the anode alternating current, the A.C. voltage between control grid and cathode is approximately

$$V_g = V - j\,\omega\,L_k\,I_a$$

and thus the alternating current which flows in the grid circuit as a result of the capacity C_{kg} between cathode and control grid is

$$I_g = j\,\omega\,C_{kg}\,V_g = j\,\omega\,C_{kg}\,V + \omega^2\,L_k\,C_{kg}\,I_a.$$

As a first approximation $I_a = SV$, where S represents the slope of the valve. Thus.

$$I_g = (j\,\omega\,C_{kg} + \omega^2\,L_k\,C_{kg}\,S)\,V.$$

The first term in parenthesis represents a capacitative admittance which can be eliminated by connecting a suitable self-induction in parallel (this is done automatically in tuning the input circuit). The last term then remains and this is just the ohmic admittance $1/R_{L_k}$.

With the help of *fig. 4* we shall now study how the signal current and the fluctuation current in the anode circuit of the amplifier valve change due to the self-induction of the cathode connections.

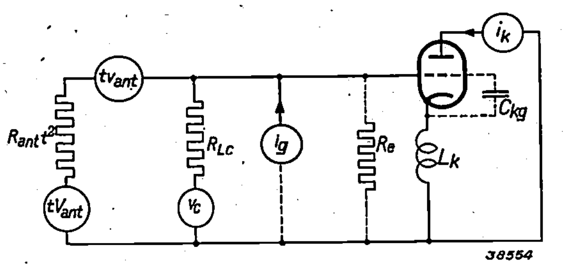


Fig. 4. Extended equivalent circuit diagram of the connections given in fig. 1. Besides the elements shown in fig. 3 the self induction L_k of the cathode connections and the capacity C_{gk} between cathode and control grid have been taken into account.

At first glance it might be supposed that it is simply necessary to replace the input resistance $[R_{ant}t^2|R_{LC}|R_e]$ between control grid and cathode by the smaller value $[R_{ant}t^2|R_{LC}|R_i]$ everywhere in the formulae. Since I_a as well as $\delta_1 I_a$, $\delta_2 I_a$ and $\delta_3 I_a$ change proportionally with the input resistance, the quotient $\delta I_a/I_a$ would not be affected by this change, as far as the first three sources of fluctuation are concerned. The fourth source of fluctuation $\delta_4 I_a$, however, appears according to (7d) to be independent of the input resistance, so that the relative intensity of these fluctuations would have to increase with decreasing input resistance. Further considerations which will be given below, will, however, show that this conclusion is incorrect; the relative intensity of these last fluctuations also remains quite unchanged. The self-induction of the cathode connections of an amplifier valve thus has no effect on the ratio between fluctuation current and signal current, so that equation (8) need not be corrected for this.

The decrease of $\delta_4 I_a$ with increasing damping due to the self-induction of the cathode connections may be explained as follows. A spontaneous increase in the anode current causes a change in the cathode potential due to the self-induction of the cathode connections. Due to the capacity between control grid and cathode, this change causes the flow of grid currents, which in turn cause the control grid potential to change, in the sense that it decreases. Hereby the spontaneous increase of the anode current is partially suppressed; the remaining effect is found to be given by the formula:

$$\delta_4 I_a = i_k \frac{[R_{ant}\, t^2|R_{LC}|R_i]}{[R_{ant}\, t^2,R_{LC}|R_e]}, \quad \ldots \ldots \quad (11)$$

so that $\delta_3 I_a$ actually does change proportionally with the input resistance $[R_{ant}t^2|R_{LC}|R_i]$.

In order to confirm equation (11) let us consider the equivalent circuit given in *fig. 5a*. Due to the anode current fluctuation $\delta_4 I_a$, a voltage $j\omega L_k \delta_4 I_a$ occurs over the self-induction L_k. Since the impedance of the capacity C_{kg} connected in series with this is still large, even at very high frequencies, compared with that of L_k, we may consider L_k in the circuit sketched as a source of voltage without internal resistance. By again applying the theorem of fig. 3 to the branch L_k-C_{kg}, we transform

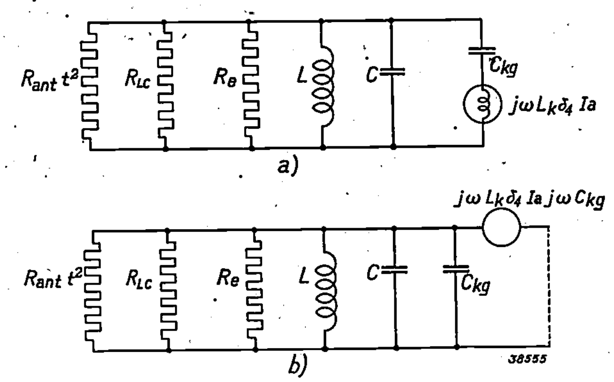


Fig. 5. Diagram showing the reaction of the anode circuit on the control grid circuit *via* the self-induction L_k and the capacity C_{kg}. Diagram *b*) is derived from *a*) by applying the theorem of fig. 3 to the branch L_k-C_{kg}.

the voltage source into a current source again (see fig. 5*b*). The current from the source is $j\omega L_k \delta_4 I_a \cdot j\omega C_{kg}$. The current is divided among the branches $R_{ant}t^2$, R_{LC}, R_e, L, C, C_{kg}. We shall now consider only those components of the current fluctuations with frequencies in the neighbourhood of the resonance frequency. For these the following is valid:

$$\omega^2 L\,(C + C_{kg}) = 1$$

In the neighbourhood of this frequency the currents through the reactive branches L_1, C, C_{kg} practically cancel each other, so that the whole current from the source flows through the resistance branches. This current therefore causes a grid voltage variation (voltage across C_{kg}) of the magnitude:

$$\delta_4 V_g = j\omega L_k\, \delta_4\, I_a\, j\omega C_{kg}\, [R_{ant}\, t^2|R_{LC}|R_e].$$

The result of this grid voltage variation is an anode current fluctuation $\delta_4 V_g \cdot S$. By adding to this the anode current fluctuation i_k we obtain the total anode current fluctuation $\delta_4 I_a$, which is thus given by the relation:

$$\delta_4 I_a = -\omega^2\, L_k\, C_{kg}\, S\, [R_{ant}\, t^2|R_{LC}|R_e]\, \delta_4 I_a + i_k$$

This equation is easily solved for $\delta_4 I_a$ and one obtains

$$\delta_4 I_a = \frac{i_k}{[R_{ant}\, t^2|R_{LC}|R_e]\left(\dfrac{1}{[R_{ant}\, t^2|R_{LC}|R_e]} + \omega^2 L_k\, C_{kg}\, S\right)}.$$

If, finally, we take into account the fact that $\omega^2 L_k C_{kg} S = 1/R_{Lk}$ (equation 10), we may simplify the factor in parenthesis in the following way:

$$\frac{1}{[R_{ant}\, t^2|R_{LC}|R_e]} + \frac{1}{R_{Lk}} = \frac{1}{[R_{ant}\, t^2|R_{LC}|R_i]}$$

and we thus obtain:

$$\delta_4 I_a = i_k \frac{[R_{ant}\, t^2|R_{LC}|R_i]}{[R_{ant}\, t^2|R_{LC}|R_e]},$$

which confirms equation (11).

Discussion of the result

Equation (8) gives the ratio between the fluctuation current and the signal current in the anode circuit of the first amplifier valve as a function of the signal voltage and of the various fluctuation voltages and currents. Now the fluctuation voltages and currents themselves are not known; the theoretical formulae which were given at the beginning of this article only give the mean squares of these fluctuations. Therefore we can only use equation (8) when both sides are squared and the mean value of that power is considered. If we work out the third term in braces and consider also that the mean values of the product of two different fluctuations are equal to zero [7]) we obtain:

$$\frac{\overline{\delta I_a^2}}{I_a^2} = \frac{\overline{v_{ant}^2}}{V_{ant}^2} + \frac{R_{ant}^2\, t^2}{V_{ant}^2}\left\{\frac{\overline{v_C^2}}{R_{LC}^2} + \overline{i_g^2} + \frac{\overline{i_k^2}}{S^2}\left(\frac{1}{R_{ant}\, t^2} + \frac{1}{R_{LC}} + \frac{1}{R_e}\right)^2\right\} \quad \ldots \ldots \quad (12)$$

The values $\overline{v_{ant}^2}$, $\overline{v_C^2}$, $\overline{i_g^2}$ and $\overline{i_k^2}$ have already been given in the beginning of this article by the formulae (1), (2), 3c) and (5). With these formulae we obtain:

$$\frac{\overline{\delta I_a^2}}{I_a^2} = 4\,k\,T\,\Delta f\,\frac{R_{ant}^2\, t^2}{V_{ant}^2}\left[\frac{a}{R_{ant}\, t^2} + \frac{1}{R_{LC}} + \frac{5{,}6}{R_e} + R_r\left(\frac{1}{R_{ant}\, t^2} + \frac{1}{R_{LC}} + \frac{1}{R_e}\right)^2\right] \quad \ldots \quad (13)$$

In this expression the relation between fluctuation current and signal current is represented as a function of aerial properties, circuit properties and valve properties.

A quantity which has not been discussed until now is the transformation ratio t. There are various points of view to be considered in the choice of this ratio.

In the case of the ordinary broadcasting receiving sets it is considered particularly important that the resistance and the capacity of the aerial should not cause too great a detuning of the input circuit, so that a set with a calibrated scale can be connected to different aerials. For this reason the transformation ratio t is chosen very large. It may be seen from equation (6) that the signal voltage on the control grid approaches zero for very large as well as for very small values of the transformation ratio; increasing t beyond a certain limit will therefore be impossible without a loss of selectivity.

If it is desired to be able to receive very weak signals, it is obvious that t must be chosen so that the sensitivity is as great as possible. The condition for this, derived from equation (6) is the following:

$$1/t^2 = R_{ant}\left(\frac{1}{R_{LC}} + \frac{1}{R_i}\right) \quad \ldots \ldots \quad (14a)$$

With this choice of t the largest signal is obtained at the grid of the amplifier valve. This does not, however, immediately mean that the quality of the reception with this choice of t will be the best. As mentioned in the foregoing the requirement of the best possible reception leads to the condition that the signal amplitude I_a should not be as large as possible in absolute value, but in relation to the fluctuation amplitude δI_a. If from equation (13) the value of t is calculated for which this condition is fulfilled, one finds that

$$1/t^2 = R_{ant}\sqrt{\left(\frac{1}{R_{LC}} + \frac{1}{R_e}\right)^2 + \frac{1}{R_r}\left(\frac{1}{R_{LC}} + \frac{5{,}6}{R_e}\right)} \quad \ldots \ldots \quad (14b)$$

i.e. a relation which will often furnish a considerably smaller value of the transformation ratio. If t is chosen according to equation (14) the following relation is obtained for the ratio between fluctuation current and aerial current:

$$\frac{\delta I_a^2}{I_a^2} = \frac{4\,kT\,\Delta f}{V_{ant}^2/2R_{ant}}\left[\frac{a}{2} + R_r\left(\frac{1}{R_{LC}} + \frac{1}{R_e}\right) + \sqrt{\frac{R_r}{R_{LC}} + 5{,}6\,\frac{R_r}{R_e} + R_r^2\left(\frac{1}{R_{LC}} + \frac{1}{R_e}\right)^2}\,\right] \quad \ldots \quad (15)$$

For a given signal energy and given values R_r, R_{LC} and R_e this formula gives the optimum ratio between fluctuation current and signal current. We shall now consider two limiting cases, namely the case of normal broadcasting frequencies and the case of very high frequencies.

At *broadcasting frequencies* R_e amounts to many megohms for ordinary amplifier valves, while R_r is in

[7]) This is always the case when two chance fluctuations are statistically independent of each other. In the case of the fluctuations i_g and i_k this condition is not fulfilled. Nevertheless the average of the product it disappears, since the two fluctuations have a relative phase difference of practically 90°.

general only several hundred ohms (see equation (4)). The contribution of the electron input damping to the noise may then be entirely neglected.

If we neglect the cosmic noise, which does not actually represent any lack of perfection of the receiver, but a lack of perfection in the incoming signal, the a between the brackets disappears. The ratio between fluctuation current and signal current at low frequencies is then almost determined by R_r/R_{LC}, and we see that the receiver can be made suitable for the reception of extremely weak signals by choosing the circuit resistance R_{LC} sufficiently large. As an objection to this it is often put forward that the circuit resistance should not exceed a certain value, since otherwise an undesired high selectivity is obtained. This objection is, however, unjustified: the selectivity is not determined by the circuit resistance R_{LC}, but by the total input resistance $[R_{ant}t^2|R_{LC}|R_i]$ of the connections. Now among other things R_i contains the damping due to the self-induction of the cathode connections. As we have seen, this contribution to the damping has no influence on the ratio of fluctuation current to signal current; we may therefore reduce an undesired high value of the total input resistance to the desired value with the help of this damping term. In this way we can obtain a receiving circuit which amplifies an extremely weak input signal almost free of noise at not too high frequencies.

At *very high frequencies* ($<$10 m), in addition to the circuit damping the electron input damping is also very important. With waves of the order of magnitude of 50 cm the electron input damping R_e amounts to only a few hundred ohms even for the best valves now in use, while the resistance in parallel R_{LC} of the oscillation circuit can be made almost as large as with broadcasting frequencies by the use of cavity resonators and the like. The ratio between fluctuation current and signal current is then determined according to equation (15) exclusively by R_r/R_e, thus by the ratio between noise resistance and electron input resistance of the valve. *The quotient R_r/R_e thus determines the suitability of an amplifier valve for the amplification of very weak signals of high frequency.* Use has already been made in this periodical of this result, namely in the estimation of the properties of the valves EFF 50 and EF 51 [8]).

The compensation of distribution fluctuations in pentodes [8])

In the foregoing we have assumed that a triode was used as amplifier valve. It is not, however, difficult to carry out the same calculations for valves with more grids. Since a pentode is usually used as high-frequency amplifier valve, we shall examine the fluctuation phenomena occurring in this case more closely. In a pentode the same fluctuation phenomena are present as in a triode, but in addition so-called distribution fluctuations occur due to a changing current distribution between screen grid and anode.

If the self-induction of the cathode connections is neglected, it is found that the influence of the distribution fluctuations can be taken into account by an increase of the noise resistance R_r (for instance by a factor 2), and for the rest, the formulae derived above for the triode may be retained. If, however, the self-induction of the cathode connections has a significant effect, less pleasing results are obtained. The signal currents and the fluctuation currents calculated until now are weakened in the same ratio by this self-induction. The distribution fluctuations, however, which do not pass through this self-induction, remain unchanged, and will therefore cause a relatively greater interference [9]).

A fundamentally simple manner of overcoming this objection is indicated in *fig. 6*. The screen grid

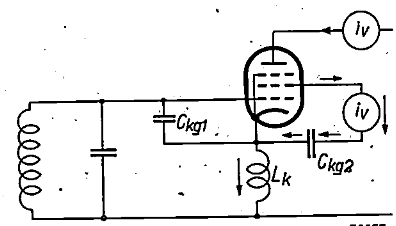


Fig. 6. By introducing a large enough capacity C_{kg2} between screen grid and cathode inside the valve, the distribution fluctuations i_v can be made to pass through the self-induction L_k of the cathode connections.

is connected to the cathode inside the tube *via* a sufficiently large capacity C_{kg2}. The screen grid

[8]) See the articles referred to in footnote [3]), as well as M. J. O. Strutt and A. van der Ziel, Physica 8, 1 and 424, 1941.

[9]) This phenomenon is more objectionable the more the signal is damped by the self-induction of the cathode connections. From this point of view the valves EFF 50 and EF 41, which have recently been described in this periodical, are very satisfactory. In the case of the valve EFF 50 the self induction of the cathode connections, thanks to the application of the push-pull principle, furnishes only a very small contribution to the input damping, while in the case of the valve EF 50, by the use of a second cathode connection, it has even been made possible that the self induction of the cathode connection does not damp but eliminates damping.

current then also flows through the cathode connections as indicated by arrows in the figure. It may be seen that now, as far as the fluctuations are concerned, the screen grid simply works as a second cathode connected in parallel with the first, so that the fluctuations i_v can be added directly to the fluctuations i_k. The result is an increase in the noise resistance R_r, but for the rest the formulae derived in the foregoing hold exactly with this altered noise resistance.

The connections indicated in fig. 6, however, do not yet represent the best method of using the pentode. The ratio between fluctuation current and signal current in the anode connections can be made still smaller if a self-induction is included not only in the cathode connections but also in the screen grid connections. We then obtain the connections shown in *fig. 7*. If for this scheme we examine the action of the signal voltage V_{ant} and of the fluctuation voltages and currents v_{ant}, v_C, i_g and i_k, we find that the currents which are excited by these sources of voltage and current in the anode circuit of a pentode can be described by exactly the same formulae as in the case of the triode, except that for R_i a slightly different value must be taken, since this input resistance is found to depend upon L_{g2} as well as L_k.

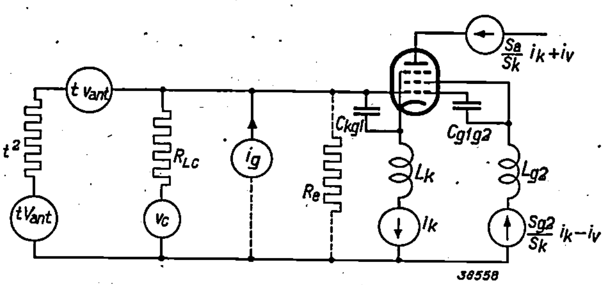


Fig. 7. Substitute diagram for an input circuit with pentode. Besides the self induction L_k of the cathode connections and the capacity C_{kg} between cathode and control grid, the self induction L_{g2} of the screen grid connections and the capacity C_{g1g2} between control grid and screen grid affect the properties of the circuit.

We may therefore say that the properties of a pentode with relation to the aerial signal and the sources of fluctuation v_{ant}, v_C, i_g and i_k do not differ appreciably from those of a triode. The great difference lies, however, in the fact that in the pentode there is a fifth source of fluctuation, namely the above-mentioned distribution fluctuations i_v.

What is the result of this on the fluctuations in the anode circuit? When there is no self-induction in the screen grid connections the distribution fluctuations must simply be added to the other fluctuations of the anode current. If, however, there is a self-induction in the screen grid connections, an attenuation of the distribution fluctuations occurs in the anode circuit due to a mechanism entirely analogous to that which is responsible for the attenuation of the cathode current fluctuations. We shall not therefore go more deeply into the calculation, but state only that a distribution fluctuation i_v in the screen grid circuit reacts on the control grid voltage and thereby leads to an anode current fluctuation:

$$i_v' = -i_v \omega^2 L_{g2} C_{g1g2} S \left[R_{ant} t^2 | R_{LC} | R_i\right] \text{ [10]}.$$

The total distribution fluctuation in the anode circuit is

$$\begin{aligned}\delta_5 I_a &= i_v + i_v' = \\ &= i_v \left(1 - \omega^2 L_{g2} C_{g1g2} S \left[R_{ant} t^2 | R_{LC} | R_i\right]\right).\end{aligned}$$

It is therefore clear that the distribution fluctuations in the anode circuit can be made equal to zero by choosing a suitable value for the self-induction L_{g2} of the screen grid connection.

If the distribution fluctuations have been compensated, the pentode behaves exactly like a triode as far as the fluctuations are concerned.

[10]) This formula is entirely analogous to the formula for the fluctuations in the anode circuit which are excited by the cathode current fluctuations i_k. The latter may be written in the form

$$\delta_4 I_a - i_k = i_k' = -i_k \omega^2 L_k C_{kg} S[R_{ant} t^2 | R_{LC} | R_i]$$

as may easily be deduced from the formulae given for $\delta_3 I_a$ in the foregoing.

A METHOD FOR THE CONTROL OF COLOUR DEVIATIONS

by J. J. WENT and P. KOOLE.

535.6.08

In the case of coloured products it is often necessary to be able to reproduce a given colour exactly at some later time. For pratical purposes it is very desirable to be able to ascertain objectively whether this requirement is satisfied or to be able to indicate objectively fixed tolerances for this condition. A simple method of doing this has been worked out and is described in this article. The arrangement is intended particularly for controlling the colours of lacquers. The "colour" of a lacquer is measured by the reflection coefficient for four separate wave lengths. The practical application of the method is illustrated by means of an example.

Wherever coloured substances or articles are made the question may arise of whether the colour obtained is "correct", *i.e.* whether it corresponds exactly to what was desired. This question may be of particular importance when it is necessary at some later date to reproduce a given colour exactly, which is often the case in the dyeing of textiles, the lacquering of surfaces and the like. The system ordinarily used in the past for the control of colour consisted in keeping a dyed or lacquered sample, with which the coloured object could be compared visually. This method was usually modified to the extent that a sample of the pigment or laquer itself was kept and at the moment when the comparison took place a new test object was dyed or lacquered with the old sample.

There are two objections to this method. In the first place uncontrollable colour changes may occur during the time the sample is kept. In the second place visual comparison involves the fact that only a subjective judgement can be obtained about the presence of any colour deviations and whether or not they may be permissible. Difficulties often arose between manufacturers and consumers due to this lack of objectively fixed and controllable colour tolerances.

Particularly for the second example mentioned above, lacquered surfaces, a simple method has been worked out in this laboratory which permits the determination of the colour by an objective measurement. This method will be described.

Principle of the method

The colour impression of a surface, for instance of a lacquered plate, is due to the fact that light of different wave lengths is reflected in different intensities. From this it follows directly that the colour impression will depend, among other factors, upon the spectral distribution of the light with which the surface is irradiated. If this spectral distribution is known, and if the reflection coefficient R of the plate is known for all visible wave lengths λ, the composition and thus also the colour of the reflected light can be calculated [1]).

In our case it is not actually a question of the colour itself, but of similarity of colour between two surfaces, no matter by what light source they are both irradiated. In order to ensure this similarity the function $R(\lambda)$ will clearly have to be the same for the two surfaces. The measurement of the "colour" of lacquers which we are here considering thus reduces to the determination of the function $E(\lambda)$[2]).

Two surfaces with the same value of $R(\lambda)$ may not actually have the same appearance. Besides the dependence on wave length the dependence on angle of the reflective coefficient will play a part. If a glossy and a dull surface are compared, the former will reflect especially strongly in those directions in which the angle of observation α is equal to the angle of incidence β of the light, while for a dull surface there is a more uniform distribution over all angles α. If therefore absolute similarity of two surfaces is required, not only must the functions $R(\lambda)$ correspond for a given angle of incidence and observation, but in general the functions $R(\lambda, \alpha, \beta)$ must correspond.

The variation of the function $R(\lambda, \alpha, \beta)$ with the angles α and β depends not only on the kind of lacquer, but to a very high degree upon the nature of the surface to which the lacquer is applied, as well as upon the way in which it is applied (sprayed, brushed, dried quickly or slowly, etc.). In order to eliminate these factors the lacquer surface to be

1) Besides the colour impression, the reflection coefficients also determine the brightness impression which is just as important in practice. For the first factor it is mainly the shape of the curve $R(\lambda)$, which is decisive, for the second, the size of $R(\lambda)$. For the sake of simplicity we speak here only of the "colour"; at the end of the article the two concepts will be separated again.

2) The determination of a colour by a point in the colour triangle, as has been described and applied in this periodical (e.g. P. J. Bouma, Philips techn. Rev. 1, 282, 1936 and recently by A. A. Kruithof, Philips techn. Rev. 6, 65, 1941) would not be sufficient here, because a given surface may have very different colour points according to the nature of the illuminating source.

investigated was applied to a glass plate and the reflection coefficient of the lacquer surface against the glass was measured, see *fig. 1*. Because of the good reproducibility of this kind of surface it may in this case be assumed that with the same kind of lacquers the $R(\lambda, \alpha, \beta)$ functions will have the same character as far as their dependence on α and β is concerned, and will only differ in absolute size

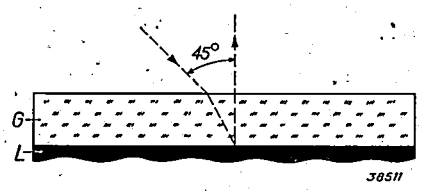


Fig. 1. A layer of lacquer L is applied to a glass plate G. The reflection of the lacquer surface against the glass is investigated for the angles of incidence and reflection indicated. Taking the refraction in the glass into account, the light must be incident on the plate at an angle of about 45°.

(*i.e.* in the total reflection coefficient). It is therefore possible in measuring R to confine oneself to a single angle of incidence and a single angle of observation. For the former an angle of 30° was chosen, while the observation took place approximately normal to the surface.

The use of glass test plates has also a second advantage. When a layer of lacquer does not "cover" because it has not been made thick enough, the reflectivity of the underlayer will also play a part, and the value measured will not be a true property of the laquer alone. It cannot be seen directly whether or not a layer of lacquer covers entirely unless it is applied in the manner described to a transparent underlayer, since when this is done a covering layer of lacquer will transmit practically no light.

Although it is now only necessary to measure the reflection coefficients in a single definite arrangement, the determination of the complete function $R(\lambda)$ would still require a large number of measurements. As a further simplification, therefore, the reflection coefficient is determined for only four definite wave lengths which are distributed as favourably as possible throughout the whole visible region of the spectrum.

There may of course be dangers connected with this simplification. If one is concerned with the two reflection curves drawn in *fig. 2* for example and if the reflection coefficient is measured at exactly the four wave lengths λ_1-λ_4 where the curves intersect each other, the surfaces will be judged to be identical, while they will certainly not give the same colour impression under all conditions of illumination[3]). In practice, however, this objection is not so serious. In the first place the reflection curves of the most common pigments have a smooth flowing shape like curve *1* in fig. 2, and only a few

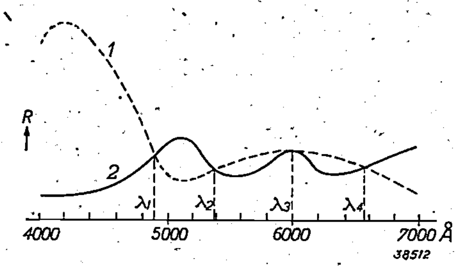


Fig. 2. A case may occur in which the curves $R(\lambda)$ of two different surfaces cut or touch each other at four points (wave lengths λ_1-λ_4).

pigments such as occur in oil paintings are observed to have several maxima and minima (curve *2* in fig. 2). The fact that two reflection curves under examination should have four points of intersection is therefore already very improbable, and that these four points of intersection should correspond exactly to the four wave lengths measured would be extremely improbable. In the second place the problem here is the comparison of approximately similar surfaces which will in principle contain the same pigments, and which therefore must exhibit qualitatively similar behaviour in their reflection curves, see *fig. 3*.

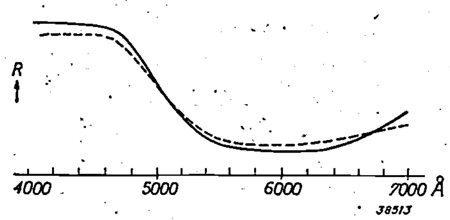


Fig. 3. Ordinary kinds of paints and lacquers have fairly smooth $R(\lambda)$ curves. The curves of two lacquers to be compared will not in general have more than two, at the most three, common points, as shown in the figure.

Description of the measuring arrangement

For the measurement of the reflection coefficient according to the method outlined four kinds of monochromatic light distributed over the whole spectrum are required. The following were chosen: in the blue, green and yellow the mercury lines 4 358, 5 461 and 5 780 Å, respectively, while for

[3]) The danger becomes less when, instead of taking four separate wave lengths, the whole spectrum is divided into four connecting wave length regions and the measurement is carried out in these regions. Such an apparatus which works with eight instead of four "blocks" is the block photometer described in this periodical (P. M. v. Alphen, Philips techn. Rev. 4, 66, 1939). The arrangement would however, become much more complicated and expensive if this principle were applied here.

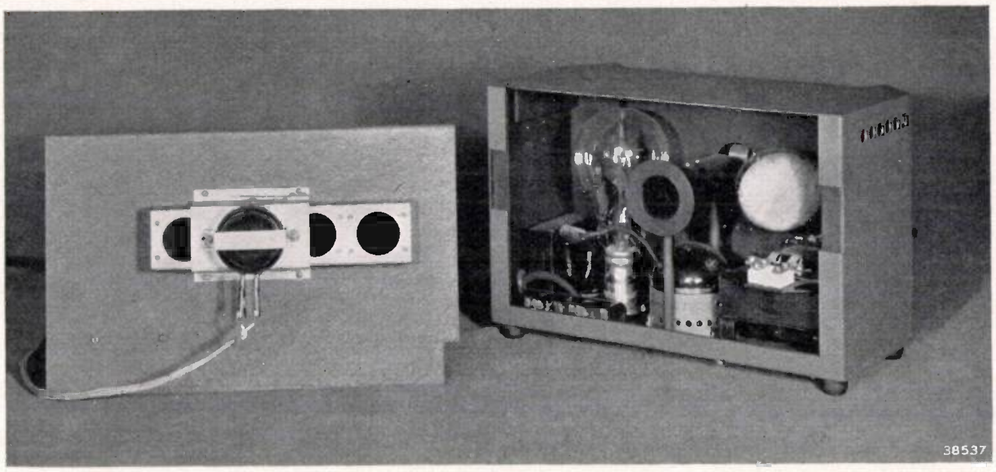


Fig. 4. The cabinet with the measuring arrangement with its front wall removed and placed beside it on the left. To the left inside the cabinet may be seen the mercury lamp, to the right the ordinary electric lamp (with half of the bulb silvered), in addition the series apparatus needed for the mercury lamp. In the middle is a tube containing a lens and below it a fan. On the outside of the front wall is a slide. with four filters and in the middle a selenium photocell.

the red, in which it is dificult to obtain a monochromatic radiation of sufficient intensity, a spectral band was used which cuts off sharply at one edge at 6 000 Å and falls off at the other edge at about 7 000 Å. As sources of light an ordinary high-pressure mercury lamp (HP 300 Dlm) and for the red an ordinary gas-filled electric lamp of 500 W were used. In the photograph *fig. 4* the two lamps may be seen to the right and left in the cabinet whose front wall has been removed. The plate to be investigated is pressed by means of a clamp against a circular opening in the rear wall of the cabinet (see *fig. 5*). The lamps are so placed that the plate is illuminated at an angle of about 45°, and due to the refraction in the glass the angle of incidence on the laquer surface amounts to about 30°. The part of the light reflected in about the direction of the normal to the surface is concentrated by a lens on a photoelement (selenium blocking-layer cell) which is set up against the front wall (see *fig. 6*). A tube shields the photocell from other directions. With the help of a slide between lens and cell different filter combinations can be brought into the path of the light. These combinations are so chosen that in each case only one of the three desired spectral lines or the spectral band in the red is transmitted.

Since, after reflection at the test plate and filtering, only a fraction of the original light flux reaches the photocell, in order to obtain easily measurable photocurrents the original light flux must be adequate. On the other hand the light flux must not be too large because the method of measurement to be described in the following makes it

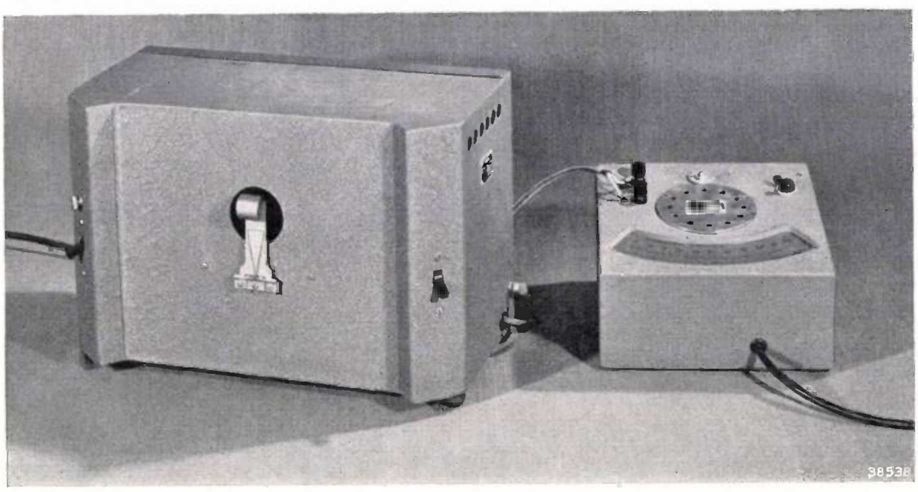


Fig. 5. Rear view of the cabinet. The circular opening against which the test plate can be pressed is visible. With the two switches to the left and right the mercury lamp and the electric lamp, respectively, can be switched on. To the right a microammeter.

desirable to have a linear relation between the photocurrent and the light flux incident on the cell. With the lamps mentioned above this linearity is ensured for all spectral regions and reflectivities of practical importance, while the photocurrents are still large enough to be measured directly with a microammeter, *i.e.* without any amplifier.

In order to avoid the development of too much heat inside the cabinet while the lamps are burning, a small fan has been built in. This may be seen in fig. 4 below the tube. It is switched on at the same time as the lamps.

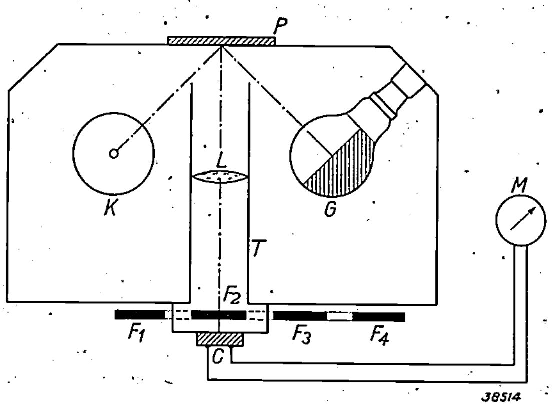


Fig. 6. Path of the rays in the measuring arrangement. K mercury lamp, G electric lamp, P test plate, L lens, T tube, F_1-F_3 filters, C selenium photocell, M microammeter.

Performance of the measurements

In order to determine the absolute value of the reflection coefficients, the light intensity of the lamps, the transmission of the filters, the sensitivity of the cell, etc. would have to be taken into account as functions of the wave length. Since, however, it is here only a question of the comparison of reflection coeficients, this is unnecessary. It is enough to compare the photo-currents obtained with the test surface or with a given standard surface. If the relation between photocurrent and light flux incident on the cell is linear, the ratio between the meter readings obtained with two surfaces is equal to the ratio between the two reflection coefficients. The required reflection coefficient is thus found in this way in per cent of the reflection coefficient of the standard surface.

In the first instance it is a matter of indifference what is used as standard surface, except that it must be satisfactorily reproducible as far as its $R(\lambda, \alpha, \beta)$ function is concerned, and it must remain constant indefinitely. In order to compare results obtained at different places with different apparatus a standard is in any case needed which satisfies the first requirement (primary standard). A very suitable surface for this purpose is that obtained by burning magnesium and allowing the magnesium oxide formed to be deposited on a plate. When sufficiently thick this surface reflects 95 per cent of the incident light very nearly according to Lambert's law (ideal diffuse reflection). In order to judge whether the layer is thick enough a mirror is used as foundation: it is only necessary to continue the deposition of magnesium oxide until no specular reflection can be observed.

The surface so obtained is easily reproducible, but at the same time very delicate. For the measurements proper, therefore, a more durable secondary standard will be used, for instance a frosted milk glass plate which can be calibrated with the primary standard in order to relate the results with different apparatus to each other. The numerical values of the relative reflection coefficients mentioned in the example below have been referred in this way to the magnesium-oxide standard.

Example of application

By the measurement of the relative reflection coefficients for four reference wave lengths we can now ascertain whether or not two lacquer surfaces are similar. If they are not similar, and there will practically always be small differences, the question immediately arises as to how large these differences may be before this actually becomes observable as a changed impression of colour or brightness. A definite (objective) answer to this question cannot be given, since only a subjective judgement is possible in the decision whether two surfaces differ visibly or not. If, however, two given surfaces have once been qualified as sufficiently alike, by common agreement between producer and consumer, and others as too different, then with the help of the arrangement here described the differences in reflection coefficients which correspond to the one case and to the other can be determined, and thereby the agreed colour tolerances are fixed for later use, and are always available.

As an example we shall discuss the investigation of a certain light grey lacquer which is used for lacquering transmitter panels and the like. This lacquer is prepared by mixing four basic pigments, namely dull white, dull black, dull red-brown and blue. The relative reflection coefficients of the original sample of the lacquer amount (in per cent of the reflection of magnesium oxide) to: for the red reference wave length $r = 26.5$ per cent, for the yellow $g = 30.5$ per cent, for the green $gr = 32.5$ per cent and for the blue $b = 36.1$ per cent.

An investigation was made of the way in which the reflection coefficients for the four reference wave lengths change when definite additional amounts of one or more of the basic pigments are added to the light grey lacquer. The difference found between the per cent values of the original and the new reflection coefficient (ΔR) is plotted in *fig.* 7. for the four reference wave-lengths (r, g, gr, b) against the per cents of extra pigment added. In fig. 7*a* white pigment was added and the reflection increased; in figs. 7*b*, *c d*, black, blue and red-brown pigment, respectively, was added, and the reflection decreased. In fig. 7*e* black and white were finally added at the same time (in a ratio of 4 to 1).

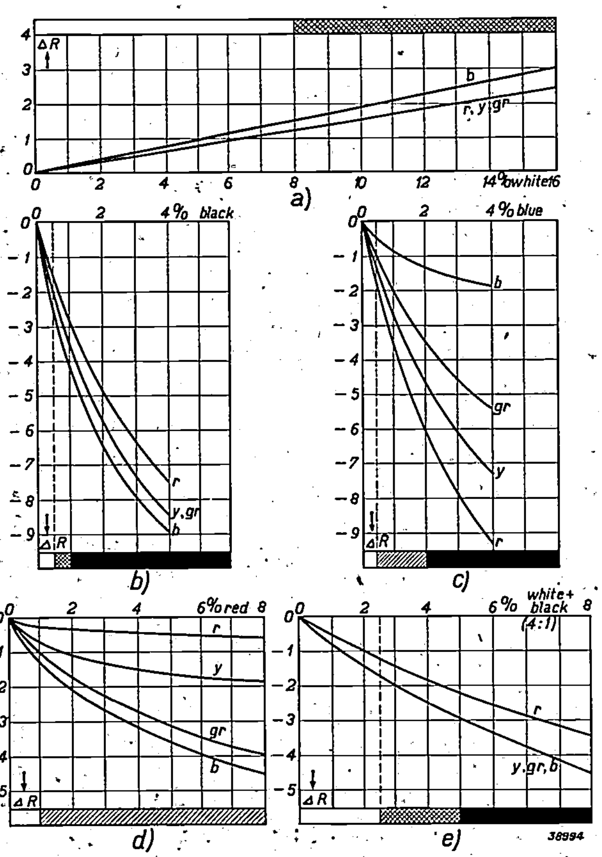


Fig. 7. In the original sample of a light grey lacquer the following reflection coefficients were found for the four wave lengths (in percent of the analogous coefficient for a magnesium oxide surface): $r = 26.5\%$, $g = 30.5\%$, $gr = 32.5\%$, $b = 36.1\%$. The change ΔR is plotted which is experienced by these values when one (or more) of the four component pigments of the lacquer is added in excess: *a*) addition of white, *b*) of black, *c*) of blue, *d*) of red, *e*) of white and black in the ratio 4 : 1. The shaded band with each figure indicates whether or not the colour and brightness impressions differ visibly from those for the original lacquer. At the percentages in the white part of the band colour and brightness are correct, in the simply shaded part only the colour differs, in the doubly shaded part, only the brightness, in the black part both factors differ.

At the same time during the measurements an estimation of colour impression and brightness of the lacquer surface was given. The question was answered (subjectively of course) of whether or not the colour and brightness deviate visibly from those of the surface with the original lacquer. In figs. 7*a-e* this estimation is represented by the variously shaded band accompanying each figure: in the region of percentages where the band is left white it is impossible to observe either difference in colour or difference in brightness compared with the original lacquer. In the part shaded in one direction only, only the colour deviates; in the doubly shaded part only the brightness, finally in the black part both colour and brightness differ.

If, without bothering about the percentages of pigment, one considers more closely the relation between the measured values of ΔR and the estimation of colour and brightness the following conclusions are reached. The colour impression is still "correct" (*i.e.* the same as that of the original lacquer) as long as the four values of ΔR for r, g, gr and b all lie within an interval of 0.8. The brightness impression is still "correct" as long as at least one of the four values of ΔR for r, g, gr or b lies below $\Delta R = 1.2$. If the subjective estimation given here is accepted by all concerned, the values may then be considered as the tolerances for the light grey lacquer in question [4]). If, for example, with a newly prepared portion of the mixture the relative reflection coefficients $r = 25.5$ per cent, $g = 30.5$ per cent, $gr = 33.5$ per cent and $b = 36.6$ per cent are measured, then the four values of ΔR are -1.0, 0, $+1.0$ and $+1.5$ respectively. The smallest of these is $\Delta R = 0$ (namely for g), the brightness impression will therefore be correct; the scattering of the values of ΔR is, however, 2.5 (namely between -1.0 for r and $+1.5$ for b), so that the colour deviation to be expected cannot be considered permissible.

In such a case it may also be deduced from the graphs how the error can be corrected. Further consideration, namely of the curves in fig. 7*e*, which were obtained by the addition of a mixture of white and black pigment, shows that the differences ΔR can be calculated in good approximation simply by adding the values of ΔR caused by each pigment separately. On the basis of the curves of figs. 7*a-d*, therefore, it can be ascertained directly what reflection coefficients will be obtained upon addition of pigment to the lacquer which has been found unsuitable. In the case described, from the too large value of b and the too small value of r

[4]) For lacquers of other colours quite different tolerance requirements may be reached, so that each case should be investigated separately.

[5]) In the mixing of pigments there is no question of an additivity for the different colours. This does not alter the fact, however, that with very small colour differences the mixing law here observed may hold for the reflection coefficients.

it could immediately be concluded that there was too little red pigment in the mixture. From fig. 7d it may now be read off that an addition of for instance 4 per cent of red pigment will give the values of ΔR for r, g, gr and b of —0.4, —1.5, —2.7 and —3.1 respectively. This, added to the above-mentioned values of ΔR gives the new values: —1.4, —1.5, —1.7, —1.6. The new mixture satisfies the tolerance requirement as far as the colour impression is concerned (the scattering is only 0.3), but not the requirement as to brightness. This can now still be corrected by adding about 10 per cent of white pigment which causes all four values of ΔR to increase by about 1.7; the mixture obtained will have the values of ΔR: +0.1, 0, —0.2, +0.3, and thereby satisfy both tolerance requirements.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on applications to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabriken, Eindhoven (Holland), Kastanjelaan.

1530: C. J. Bakker: Fluctuations and electron inertia (Physica **8**, 23-43, Jan. 1941).

After having dealt with the spontaneous current fluctuations which occur in an electronic valve whose emission is limited by the space charge present in the neighbourhood of the cathode, the displacement currents are calculated which these fluctuations cause in the grids of the electronic valve (*cf.* also Philips techn. Rev. **6**, 129, 1941). These displacement currents are important in the use of electronic valves on short waves. For two different types of radio valves, namely a pentode and a four beam octode, the current fluctuations in the connection between one of the grids and the cathode are investigated experimentally as well as theoretically. The agreement between theory and experiment was found to be excellent.

1531: M. J. O. Strutt and A. van der Ziel: The results of several electron inertia effects in electronic valves I. Theoretical explanations (Physica **8**, 81-108, Jan. 1941 (Original in German).

The currents are studied which flow to the different electrodes of a vacuum tube when a small group of electrons passes from the cathode to the anode. This is further elaborated for the case where periodic current impulses leave the cathode. Under certain conditions the anode current will behave quite differently from these impulses. The same is true for a very high-frequency A.C. between control grid and cathode. This may lead to remarkable phenomena such as frequency multiplication and rectification. For a harmonic alternating voltage on the control grid the values of input resistance and amplification factor are calculated. Simple formulae are derived for the displacement currents which occur due to electrons which pass close to the conductors in radio valves. A so-called centre of gravity hypothesis is hereby formulated which makes a simple calculation possible. Finally the possibility of transition time modulation is mentioned.

1532: J. van Niekerk and M. S. C. Bliek: Het prophylactische effect van éénmaal oraal of intramusculair, toegediende groote dosis bestraald provitamine D van dierlijke oorsprong bij rachitis. (The prophylactic effect of a single large dose of irradiated provitamine D of animal origin, administered orally or intermuscularly, in the case of rickets). Ned. T. Geneesk. **85**, 522-530, Feb. 1941).

Tests have been carried out for the purpose of studying the difference in action on groups of chicks of different amounts of irradiated provitamine D of animal origin administered in a single dose either through the beak or injected into a muscle. As a criterion of the activity of the vitamine D, its effectiveness against rachitic lesions in the growing skeleton has been chosen, as observed in X-ray photographs. The bone formation of different chicks is judged individually, and the protection against rickets was determined separately as a number for each group of birds which received the same dose administered in the same way. The anti-rachitic action of the vitamine is found to depend upon the size of the single dose administered. Furthermore it was determined that an amount of vitamine D which is injected into a muscle acts more effectively and for a longer time than the same amount upon administration through the beak. Of a normal amount of vitamine D which during 12 successive weeks the chicks had consumed with their food, much too little was found to be retained to insure normal bone formation in the following weeks.

1533: J. M. Stevels: Physical properties of glasses I (Rec. Trav. chim. Pays Bas **60**, 85-86, Febr. 1941).

The way in which the specific volume of different kinds of glass depends upon the composition is discussed in this article.

1534: J. J. Went: Adsorption phenomena on massive metal surfaces measured by means of electrical contact resistances (Physica 8, Febr. 1941).

For the contents of this article the reader may refer in general to: Philips techn. Rev. **5**, 238, 1940.

VOL. 6. No. 7, p. 193-224 JULY 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

JUDGING AN AMPLIFIER BY MEANS OF THE TRANSIENT CHARACTERISTIC

by J. HAANTJES. 537.545

The properties of amplifiers can be judged on the basis of characteristics which represent the amplification of the amplitude and the shift in phase as functions of the frequency. The characteristic of the phase shift is usually left out of consideration since its meaning is not graphically expressed. In certain cases, however, the judgment of an amplifier by means of the amplitude characteristic alone leads to erroneous conclusions. The behaviour of the amplifier can then best be characterized by the so-called transient characteristic, which indicates the output signal for a discontinuous change of the input signal. In this article the amplitude characteristics and the transient characteristics are discussed for a number of common coupling networks of amplifiers. It is shown that the transient characteristic of an amplifier furnishes the same information as the amplitude and phase characteristics together. In conclusion the concept of transient characteristic is extended to high-frequency amplification where it is not a question of the entire variation of the input signal, but only of its modulation. The single sideband reception customary in television sets here leads to interesting complications which are discussed by means of a simple example.

The amplification of electrical signals is a very common problem in electrotechnology. By this amplification is meant the excitation of a voltage or current which, considered as a function of the time, gives a diagram which has the same shape but a larger amplitude than the original electrical signal.

A familiar example of the application of amplification of signals is seen in the cathode ray oscillograph. This instrument makes visible the variation with time of weak signals (for instance voltages of 1 mV), while the cathode ray tube requires a deflection voltage of about 10 V to produce a reasonable deviation. The first requirement of an amplifier which must convert the signal voltage of 1 mV into a deflection voltage of 10 V is undoubtedly that the variation with time of the deflection voltage should give a true picture of that of the signal voltage.

Another example is found in the amplifier of a radio or television receiver. The input signal of these sets is a high-frequency voltage, generally with varying amplitude. This variation of the amplitude, the so-called modulation, is the quantity to which the voltage or current generated must be proportional. The voltage or current generated when drawn as a function of the time, thus does not actually give a picture of the signal voltage itself but only of its modulation. We shall in this case also, however, speak of the amplification of the signal voltage, and we shall show that there is a far-reaching analogy between the two cases.

The object of obtaining an amplification which is accurate in shape is approximated more or less in practice, according to the circumstances, but never entirely achieved. The deviations from the desired result may be divided into two groups. The first group is formed by deformations which increase with increasing amplitude of the input signal and disappear for a sufficiently small amplitude of the input signal. In this case one speaks of distortion due to non-linear amplification. The second group of deformations is present even at an indefinitely small amplitude of the input signal, and may be ascribed to the fact that the amplified voltage at a certain instant is not given unambiguously by the magnitude of the signal voltage, but, for example, depends at the same time upon the derivative of the signal voltage with respect to time, or on an integral of the signal voltage over the time.

An important property of these deformations is that they also occur in networks which only contain elements which satisfy Ohm's law, so that each branch may be characterized by a definite resistance or a definite complex impedance. Such networks are called linear networks, and the deformation is also called linear deformation.

In this article we shall examine the way in which the properties of an amplifier, especially with respect to its linear deformation, can most suitably be characterized. It will be shown that the well-known amplitude characteristic of an amplifier, which represents the amplitude of the input signal as a function of the frequency, is often unsuitable as a means of judging the properties of an amplifier. If a complete characterisation of the properties of an amplifier is desired, then, in addition to the amplitude characteristic, the phase characteristic must also be known. In this article, however, attention will be focussed upon an entirely different method of determining the properties of an amplifier, namely by means of a curve which indicates how the amplifier reacts to a single, discontinuous change in the signal voltage. This curve, which we shall call the transient characteristic of the amplifier, can be used to advantage in exactly those cases, where the amplitude characteristic alone does not give sufficient information. On the basis of a number of examples, we shall try to give a qualitative survey of the relation between amplitude characteristic and transient characteristic.

Fourier analysis and amplitude characteristic

Any signal $v(t)$ which is applied to the input of an amplifier can be approximately represented by a Fourier series with a number of frequencies ω_n:

$$v(t) = \sum_n a_n \cos(\omega_n t + \psi_n) \quad . \quad . \quad . \quad (1)$$

With a perfectly accurate amplification the output signal $V(t)$ would simply be a factor f larger than the input signal:

$$V(t) = \sum_n f a_n \cos(\omega_n t + \psi_n).$$

If, however, linear deformation is present, each frequency ω_n may be amplified by a different factor f_n, while in addition a phase shift φ_n will in general occur, so that the output signal may be written as follows:

$$V(t) = \sum_n f_n a_n \cos(\omega_n t + \psi_n - \varphi_n), \quad . \quad . \quad (2)$$

where f_n and φ_n are functions of the frequency ω_n, but not of the amplitude a_n. $f(\omega)$ and $\varphi(\omega)$, are called the amplitude and the phase characteristics respectively, of the amplifier [1]).

If for example it is assumed that the amplification factor τ is independent of the frequency ω and that moreover φ is proportional to ω, so that $\varphi_n/\omega_n =$ = constant = f, equation (2) can be converted into:

$$V(t) = f \sum_n a_n \cos(\omega_n t + \psi_n - \omega_n \tau) =$$
$$= f \sum_n a_n \cos\{\omega_n(t-\tau) + \psi_n\} = f \cdot v(t-\tau), \quad (3)$$

or, to the same effect,

$$V(t+\tau) = f \cdot v(t) \quad . \quad . \quad . \quad . \quad (4)$$

This means therefore that the output signal is exactly the same in shape as the input signal, and that it merely lags behind it by a time τ. Such a retardation is no objection for most applications, so that the fulfilment of the conditions that:

f is independent of ω, and
φ is proportional to ω

guarantees the satisfactory performance of the amplifier.

In practice indeed an attempt is usually made to give to amplifiers, which must amplify accurately as to form, the flattest possible amplitude characteristic; too little attention, however, is sometimes paid to the phase characteristic. This is understandable to some extent, when it is kept in mind that the phase behaviour in sound amplifiers, and therefore also in radio receivers, is of practically no importance. In order to make a natural impression on the ear it is unnecessary that the sound be reproduced faithfully. It is here only a question of the correct transmission of the frequency spectrum, and the phase relation between the different components is of little importance within wide limits. This is the reason why, in addition to the absence of non-linear distortion, the approximately flat shape of the amplitude characteristic is practically the only requirement which is here made.

In the case of oscillograph amplifiers and television amplifiers, on the other hand, a flat shape of the amplitude characteristic offers no guarantee

---

[1]) If in addition to linear deformation, non-linear deformation is also present, f and φ change not only with the frequency but also with the amplitude. Moreover, frequencies occur in the output signal which are not present in the input signal, so that, taken strictly, it is no longer possible to speak of an amplitude characteristic or a phase characteristic. On the subject of the characterization of non-linear distortion see for example Philips techn. Rev. 4, 354, 1939, equations (1) and (2).

[2]) See in this connection the article by J. F. Schouten, Philips techn. Rev. 4, 167, 1939.

of satisfactory performance. Let us consider for example a simple resistance amplifier, of which the connections are indicated in *fig. 1* ($R' \gg R$). When the valve capacities may be neglected so that the

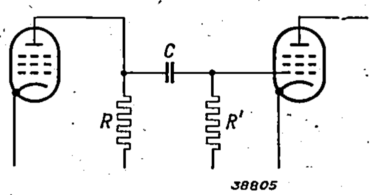


Fig. 1. Coupling between two valves of a resistance amplifier. The capacities between the electrodes of the valves are neglected in this diagram.

scheme indicated is exactly valid, the amplification factor will increase with increasing frequency and finally reach a limiting value for frequencies for which the impedance of the condenser C is small compared with R'. At these high frequencies a perfectly accurate amplification of the input is then obtained.

We now consider the reproduction of a block-shaped signal (*fig. 2a*) whose period is so short that already at the fundamental frequency 98 per cent of the maximum amplification is obtained. In this case one would be inclined to think that almost no deviation from the square form could any longer appear in the output signal. Actually, however, the variation of the output voltage obtained is that shown in fig. 2*b*, which differs very much from the block-shaped signal voltage. If the amplifier is part of an oscillograph apparatus, fig. 2*b* shows graphically the distortion of the oscillograms which must be expected.

On the basis of the amplitude characteristic of

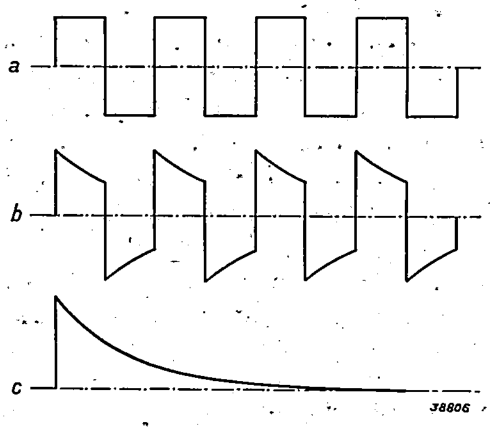


Fig. 2. a) Block-shaped input signal, by means of which the reproduction of an amplifier can be investigated. *b*) The output signal which a resistance amplifier according to fig. 1 gives with an input signal according to curve *a* when the frequency is chosen so low that the fundamental wave is attenuated by 2 per cent by the impedance of the coupling condenser. *c*) The output signal for an input signal which suddely jumps from 0 to 1 at $t = 0$ and then remains constant (transient characteristic).

the amplifier one would certainly not expect this distortion. If the phase characteristic is taken into account as well, this distortion can, however, be deduced, since amplitude characteristic and phase characteristic together give a complete picture of the behaviour of an amplifier in the absence of non-linear distortion.

Transient characteristics

As we have seen, the amplitude characteristic $f(\bar{\omega})$, which describes the reproduction of sinusoidal signals, does not furnish sufficient information about the fidelity with which a given signal is reproduced. The oscillogram of a block signal, fig. 2*b*, approaches this result much more closely, so that this oscillogram might be considered directly as a characteristic of the amplifier. This method can be still somewhat simplified by choosing as input signal instead of the series of discontinuous voltage changes of which the block signal is made up, a single discontinuous increase of the voltage, It was Heaviside who first investiged the behaviour of all kinds of electrical networks with such an input signal. The treatment of the mathematical problems thereby encountered led him to the development of operational calculus [3]). In this article, however, we shall work only with ordinary differential equations. In the case of the resistance amplifier dealt with above, upon a discontinuous increase in the input signal, the output voltage will also increase discontinuously and then decrease exponentially to zero. This "transient characteristic" of the resistance amplifier is reproduced in fig. 2*c*. From the slope of the exponential curve, which represents the variation of the output voltage for a constant input voltage, it is easy to estimate the distortion to be expected in a block signal of a given periodicity.

Examples of transient characteristics

Ordinary amplifiers are so constructed that they give a reasonably faithful reproduction for a given frequency range, while for higher as well as for lower frequencies distortions occur. The example, on the basis of which we developed the concept of transient characteristic, referred to the behaviour of amplifiers on the low-frequency side of that range. The most interesting problems occur, however, on the upper boundary of the frequency region of accurate reproduction. Repeatedly in this periodical the means have been discussed which can

---

[3]) A concise explanation of operational calculus and its application to discontinuous phenomena was given in this periodical by B. van der Pol and Th. J. Weyers, Philips techn. Rev. 1, 363, 1936.

be employed for the improvement of the accuracy of the amplification for signals with these high frequencies, not only in the case of the amplifier valves themselves but also in the networks for

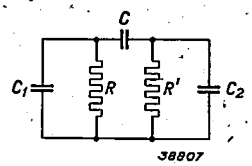


Fig. 3. Complete coupling network of a resistance amplifier. It contains the parasitic capacities C_1 and C_2 which result in a decrease in the amplification for high frequencies.

coupling the successive valves [4]). Since linear distortion originates mainly in the networks, we shall limit our considerations to them. If we again start with a resistance amplifier, an undistorted amplification can only be obtained when the anode impedance of the amplifier valve behaves as a pure ance. An attempt is here made to compensate as far as possible the fall in the total anode impedance due to the capacities C_1 end C_2 by a corresponding increase in the impedance of L. If the fundamental shortcomings of this scheme are studied, the network *III* is reached as the first stage in its improvement, etc.

An important problem is now to ascertain to what extent these coupling networks used in oscillograph amplifiers and television amplifiers actually do produce better results than the simple resistance coupling. We shall study this problem on the basis of amplitude characteristics as well as on that of transient characteristics. For the sake of simplicity we shall neglect deformations by the amplifier valves and consider only the coupling networks. The input signal is then best chosen not as a certain voltage, but as a certain current which represents the anode current of the preceding amplifier valve.

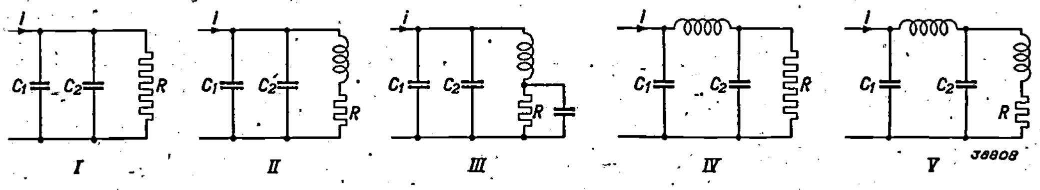


Fig. 4. Coupling network of fig. 3 and several improved coupling networks whose amplitude characteristics remain flat to higher frequencies.

resistance at all frequencies under consideration. In the scheme indicated in fig. 1 this would be true above a definite minimum frequency. Actually, however, besides the circuit elements indicated in fig. 1, there are always parasitic capacities present, namely capacities between the electrodes of the amplifier valves as well as between certain parts of the wiring and earth. If these capacities are taken into account the coupling network given in *fig. 3*, is obtained between two successive valves, and it is clear that now there can be no question of a constant anode impedance; the impedance decreases with increasing frequency as a result of the parasitic capacities C_1 and C_2.

Various connections have been devised in order to oppose the fall in impedance up to as high frequencies as possible. Four different examples of improved coupling networks are shown in *fig. 4* beside the simplest coupling network. The network *II* can for instance be derived from *I* by connecting a self induction in series with the anode resist- The output signal of the network is a certain voltage, namely the grid alternating voltage of the following valve. The calculation of the amplitude characteristics is based upon familiar principles and need not be described in detail here. There is a great diversity of possible results, since those circuit elements which are not expressly indicated by letters in fig. 4 may be chosen arbitrarily. If these elements are always chosen so that the amplitude characteristic is as flat as possible at low frequencies, then for the five networks indicated the amplitude characteristics represented in *fig. 5* are obtained. In this figure the ratio between output

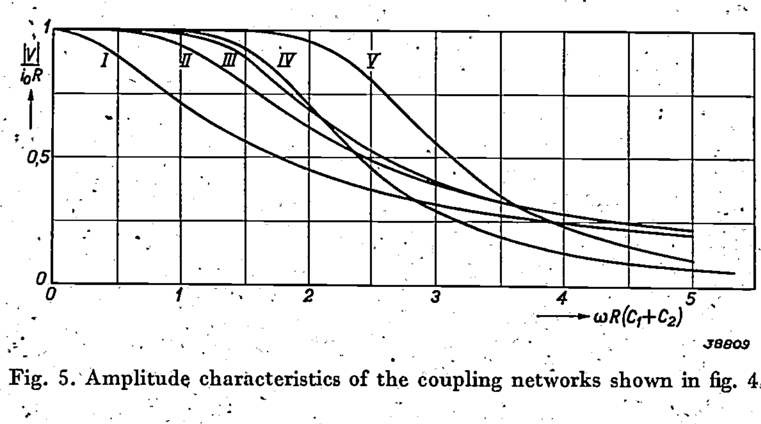


Fig. 5. Amplitude characteristics of the coupling networks shown in fig. 4.

4) Philips techn. Rev. 4, 342, 1939.

voltage and input current is plotted as a function of $\omega R (C_1 + C_2)$, where the value of this ratio is set equal to unity for very low frequencies.

On the basis of the graphs of fig. 5, one is led to the conclusion that the accuracy of the amplification at high frequencies is very much improved by the elaboration of the coupling network. If, for example, a decrease in the amplitude to 90 percent of the maximum value is considered permissible, then the upper limiting frequency of network *II* is more than double that of the pure resistance coupling *I*, while the improvement in the network *V* even amounts to more than a factor 4. On the basis of experience gained with the amplitude characteristic at low frequencies, however, we know that it is still uncertain what is the significance of these results for the accurary of the amplification of any arbitrarily varying signal.

We shall now consider the transient characteristics of the same five coupling networks. At the moment $t = 0$ the input current jumps from $i = 0$ to $i = i_0$ and keeps this value. What is required for each network is the behaviour of the voltage on the capacity C_2. It is here always assumed that up to the moment $t = 0$ there is no electrical energy present in the circuit. The condensers thus have the charge zero and the self-inductions a zero current.

As an example, the calculation will be carried out in the following for the case of the network *II*. For this purpose it is necessary to choose a definite value for the selfinduction which occurs in these connections. In order to obtain the amplitude characteristic reproduced in fig. 5, curve *II*, this self-induction must possess a value $L = 0.414\, R^2(C_1 + C_2)$. If, further, we let $C_1 + C_2 = C$, the voltages and currents in the network satisfy the following equations:

$$\left.\begin{aligned} V = \int_{-\infty}^{t} \frac{i_C}{C}\,\mathrm{d}t &= L\frac{\mathrm{d}i_L}{\mathrm{d}t} + Ri_L, \\ i_0 &= i_C + i_L. \end{aligned}\right\} \quad \ldots \quad (5)$$

The voltage V is required as a function of the time. Elimination of i_L and i_C gives:

$$LC\frac{\mathrm{d}^2V}{\mathrm{d}t^2} + RC\frac{\mathrm{d}V}{\mathrm{d}t} + V - Ri_0 = 0.$$

If $V - Ri_0 = v$ then the following holds for v:

$$LC\frac{\mathrm{d}^2v}{\mathrm{d}t^2} + RC\frac{\mathrm{d}v}{\mathrm{d}t} + v = 0.$$

By substitution of $v = A\,e^{at}$ the so-called characteristic equation is found

$$a^2 + \frac{R}{L}a + \frac{1}{LC} = 0,$$

from which it follows that

$$a = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}},$$

or, with the above indicated choice of $L = 0.414\, R^2C$:

$$a = -\frac{1{,}209}{RC} \pm j\,\frac{0{,}980}{RC}.$$

The general solution is therefore

$$v = Ae^{-1{,}209\, t/RC}\cos(0{,}980\, t/RC + \varphi)$$

or

$$V = i_0R + Ae^{-1{,}209\, t/RC}\cos(0{,}980t/RC + \varphi) \quad \ldots \quad (6)$$

A and φ must now be derived from the limiting conditions. The two initial conditions are given:

$$V(t)_{(t=0)} = 0 \quad \ldots \ldots \ldots \quad (7a)$$

$$i_L(t)_{(t=0)} = 0 \quad \ldots \ldots \ldots \quad (7b)$$

From (7b) it follows that:

$$i_C(t)_{(t=0)} = i_0$$

and since

$$i_C = C\frac{\mathrm{d}V}{\mathrm{d}t}, \quad \text{then} \quad \left(\frac{\mathrm{d}V}{\mathrm{d}t}\right)_{(t=0)} = \frac{i_0}{C}.$$

For A and φ the following equations then result:

$$i_0R + A\cos\varphi = 0,$$

$$1{,}209\cos\varphi + 0{,}980\sin\varphi + i_0R/A = 0.$$

From this one finds that

$$A = -1{,}022\, i_0R,$$

$$\varphi = -0{,}209.$$

The complete formula for the transient characteristic is thus:

$$V = i_0R\left\{1 - 1{,}022\, e^{-1{,}209\, t/RC}\cos(0{,}980t/RC - 0{,}209)\right\} \quad \ldots \quad (8)$$

The results are reproduced in *fig. 6* in which curve *II* corresponds to equation (8). As unit of voltage the value i_0R is taken, as unit of time the product $RC = R\,(C_1 + C_2)$. It is clear from the figure that a certain time elapses before the voltage reaches the final value. Furthermore, in all cases except that of the oridinary resistance coupling (curve *I*)

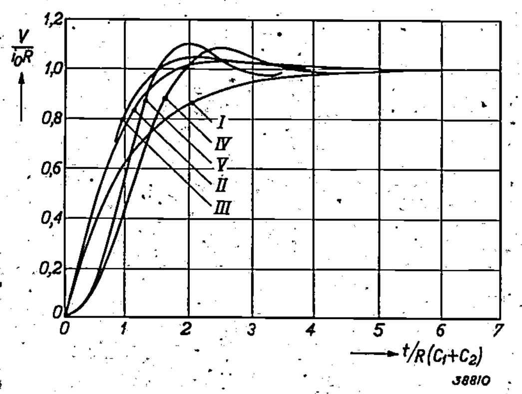


Fig. 6. Transient characteristics of the coupling networks given in fig. 4.

a certain degree of "overshooting" occurs. When the successive transient characteristics *II*, *III*, *IV*, *V* are compared, it is seen that the maximum slope in the rising part becomes steadily steeper, at the same time, however, the overshoot becomes steadily greater. It is somewhat a question of subjective judgment which of these curves should be considered as the best approach to a rectangular voltage variation. The great improvement which seems to have been obtained upon a comparison of amplitude characteristics *II* and *V* is therefore of doubtful value for the reproduction of a rectangular voltage variation.

If an attempt is made to generalize the results here found, the question arises as to what shape of the amplitude characteristic must be considered the best. The higher the frequencies to which the amplitude characteristic of an amplifier is flat, the steeper the front of its transient characteristic, which of itself is favourable for the attainment of a faithful reproduction. As we saw above, however, with increasing steepness of the front of the transient characteristic, the tendency toward overshooting also increases. This overshoot indicates the presence of weak damped resonances which will lead to a local steep slope in the the amplitude characteristic. The attempt is often made to increase the amplification of high frequencues by means of resonance in such a way that a peak occurs in the amplitude characteristic. In that case the overshoot is found to be so strong that there can no longer be any question of an improvement in the reproduction. An extension of the flat part of the amplitude characteristic to higher frequencies will therefore involve no improvement when it is accompanied by the appearance of resonance peaks or by a large increase on the slope of the descending part of the characteristic.

With this in mind, it is not surprising that connections *IV* behave less favourably than connections *III* as far as the transient characteristic is concerned. The flat part of the amplitude characteristic is but little extended upon transition from *III* to *IV*, while the slope of the descending part has increased considerably. The transition from the coupling network *I* to coupling network *II*, on the other hand, must be considered as a very appreciable improvement: in this case the flat part of the amplitude characteristic is considerably extended, without the slope of the descending part having become much greater.

Output voltage with a given input signal

In the following way it can be understood that the knowledge of the transient characteristic is sufficient to calculate the voltage variation with

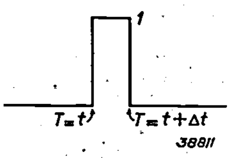


Fig. 7. Current impulse as input signal. With the help of the transient characteristic the corresponding output signal can be calculated.

any given input current. When at the moment $T = 0$ the input current jumps from 0 to 1, the voltage variation will be given by the transient characteristic $S\,(T)$. If this jump occurs at the moment $T = t$, the voltage variation is given by $S\,(T-t)$. If a current of the form indicated in *fig. 7* is taken as input current, the voltage becomes

$$\begin{aligned} V(T) &= S(T-t) - S(T-t-\Delta t) \\ &\approx S(T-t) - [S(T-t) - \Delta t S'(T-t)] \\ &= S'(T-t)\Delta t. \end{aligned}$$

Any given variable current can always be considered as a connected series of current impulses of the type represented in fig. 7 (see *fig. 8*) and having an amplitude $i(t)$.

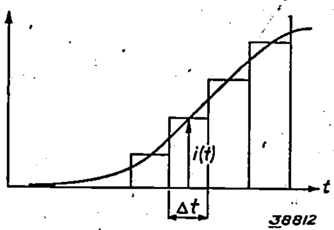


Fig. 8. Any given input signal can be built up of a number of connected current impulses.

The output voltage is obtained by adding together the contributions of all the current impulses, thus

$$V(T) = \sum_{t=0}^{t=T} i(t) S'\,(T-t)\Delta t.$$

which for $\Delta t \to 0$ passes over into

$$V(T) = \int_0^T i(t) S'(T-t)\,\mathrm{d}t \quad . \; . \; . \; . \quad (9)$$

When $i(t)$ and $S(t)$ are given the variation of the voltage may therefore be formed by integration.

As a special case we choose a current which begins at the moment $t = 0$ and is sinusoidal. The output voltage $V(T)$ is given by the formula

$$V(T) = \int_0^T \sin \omega t S'\;(T-t)\mathrm{d}t,$$

which by means of simple transformations can be converted into

$$V(T) = \sin \omega T \int_0^T \cos \omega t\, S'(t)\mathrm{d}t -$$

$$- \cos \omega T \int_0^T \sin \omega t\; S'(t)\mathrm{d}t \quad . \; . \; . \; . \quad (10)$$

The integrals in equation (10) approach certain limiting values with increasing value of T. In other words, the voltage $V(T)$ finally takes a sinusoidal form with a given amplitude and phase. If we set

$$\int_0^\infty \cos \omega t\, S'(t)\mathrm{d}t = c(\omega),$$

$$\int_0^\infty \sin \omega t\, S'(t)\mathrm{d}t = s(\omega),$$

then finally

$$V(T) = c(\omega) \sin \omega t - s(\omega) \cos \omega t$$

and from this follows the amplitude characteristic:

$$f(\omega) = \sqrt{c^2(\omega) + s^2(\omega)},$$

as well as the phase characteristic

$$\mathrm{tg}\, \varphi(\omega) = s(\omega)/c(\omega).$$

This shows therefore that the transient characteristic can replace the amplitude characteristic as well as the phase characteristic.

Transient characteristic of several stages in cascade connection

Equation (9) is of direct use when it is desired to derive the transient characteristic of an amplifier with several stages from the transient characteristics of the separate stages. Let us assume that an amplifier consists of two stages, the first of which has a coupling network with a transient characteristic S_a, and the second a coupling network with a transient characteristic S_b. If the transient $0 \rightarrow 1$ is applied to the first stage, the output voltage of that stage becomes the voltage $S_a(T)$. If the amplifier valve is a pentode, this voltage is converted

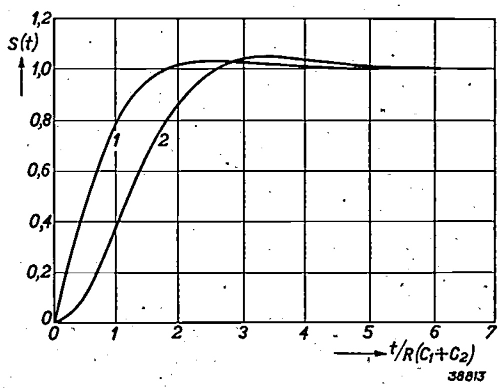


Fig. 9. 1) Transient characteristic of a stage with network *II* of fig. 4. 2) Transient characteristic of two such stages in cascade connection. The front of the characteristic becomes less steep and the peak formation more pronounced.

into a current having the same variation with time and this current is applied to the second stage having a transient characteristic $S_b(T)$. The output voltage of this second stage, which now represents the transient characteristic of the whole amplifier, can immediately be written in the following form on the basis of equation (9):

$$S(T) = \int_0^T S_a(t) S_b'(T-t)\mathrm{d}t. \quad . \quad . \quad . \quad (11)$$

By this "multiplication", therefore, the transient characteristic of an amplifier can be derived from those of the separate stages [5]).

As an example the transient characteristic of two similar stages in cascade connection was calculated. A characteristic for each stage like curve *II* of fig. 6 was assumed. In *fig. 9* the resulting transient characteristic is shown compared with that of one separate stage. The transient characteristic of the two-stage amplifier is somewhat less steep and shows a greater degree of peak formation. When still more stages are used this effect will in general appear even more pronounced.

High-frequency amplifiers

In the high-frequency stages of radio receivers and television receivers one is not concerned with the shape of the whole curve of the output signal, but only with the variation with time of its amplitude. It is therefore reasonable in the case of these stages to indicate as transient characteristic the variation of the amplitude of the output signal for a high frequency input signal whose amplitude jumps from 0 to 1. When the carrier-wave frequency of the output signal is high compared with the highest modulation frequencies, it is found to be capable of being represented approximately as a high-frequency alternating voltage modulated in amplitude, with the same frequency and phase as the input signal. This results in the fact that the modulation of the output signal for any given variation of the input amplitude can be derived from the above defined transient characteristic. As in the foregoing the input signal can be separated into the alternating currents during successive time intervals Δt, each of these "A.C. impulses" gives at the output side a voltage contribution of the same frequency and phase, so that the total output

[5]) The integral given is a symmetrical function of S_a and S_b, so that the transient characteristic does not change when the order of the stages is reversed. This is shown by a partial integration:

$$\int_0^T S_a(t)\, S_b'(T-t)\, \mathrm{d}t = S_a(0)S_b(T) - S_a(T)S_b(0) + \int_0^T S_a'(t)\, S_b(T-t)\, \mathrm{d}t.$$

The first two terms on the right hand side disappear since $S_a(0) = S_b(0) = 0$. If in the third term we set $T-t = t'$, the whole equation assumes the form

$$\int_0^T S_a(t)\, S_b'(T-t)\mathrm{d}t = \int_0^T S_b(t')\, S_a'(T-t')\, \mathrm{d}t',$$

which clearly shows the symmetry.

amplitude at every moment is given by the sum of the amplitudes of all the voltage contributions.

In an earlier article in this periodical it was pointed out that many coupling networks of high frequency amplifier stages can be derived by a simple transformation from the coupling networks of a video-frequency amplifier [6]). It is only necessary to time each capacity with a self-induction connected in parallel with a definite frequency f_0 in the middle of the television band, and bring every self-induction into resonance at that same frequency by means of a series condenser.

If one compares the transient characteristics of a video-frequency coupling network and of a high-frequency coupling network derived from it, it is found that they both possess exactly the same shape, but that the latter varies twice as slowly with time as the former. This corresponds entirely with the property of the high-frequency network derived in the above mentioned article, that its amplitude characteristic plotted as a function of $2(f-f_0)$ has the same variation as the amplitude characteristic of the video-frequency coupling network plotted as a function of f. A proof of this statement may therefore be omitted.

High-frequency amplification of carrier wave and one side band

The high-frequency coupling networks introduced in the above with branches which are tuned to a frequency f_0 possess an amplitude characteristic symmetrical on both sides of this frequency, so that two signals with the frequencies $f_1 = f_0 + \Delta f$ and $f_2 = f_0 - \Delta f$ are amplified to the same degree. If f_0 is chosen equal to the carrier-wave frequency, therefore, the two side bands will furnish the same contribution to the output signal.

At the present time it is often recommended that the receivers should be so arranged that in addition to the carrier wave they will receive only one side band. This is chiefly because then with a given width of the frequency band for which the amplifier is sensitive twice as high modulation frequencies can be amplified. The carrier wave then lies at the upper or lower limit of the frequency band in question, and is usually displaced so much to one side that the amplication for the carrier wave has already fallen to one half. We shall now discuss briefly the results which are thereby obtained for the case of one stage which contains a simple circuit with damping in parallel.

In the very first place it is found that it is impossible with such an amplifier to derive the behaviour of the output amplitude for any given behaviour of the input amplitude from the transient characteristic. This is connected with the fact that the output signal of this amplifier does not exhibit pure amplitude modulation upon a sudden change in amplitude of the input signal but also varies in phase and frequency. The first reaction to a discontinuous change in the input signal is the appearance of a free oscillation with a frequency lying in the middle of the band to which the amplifier is sensitive. This oscillation is gradually damped: only the forced oscillation then remains whose frequency, as stated above, lies at the edge of the sensitive region.

The result of this frequency modulation is that the contributions of the different preceding A.C. impulses to the output voltage present at a definite single instant differ mutually in frequency. The amplitude of the output voltage is then no longer given by the sum of the amplitudes of all the contributions, so that the variation of the output voltage can no longer be calculated by integration over the contributions of all the preceding A.C. impulses, but it must be treated separately for each variation of the input current.

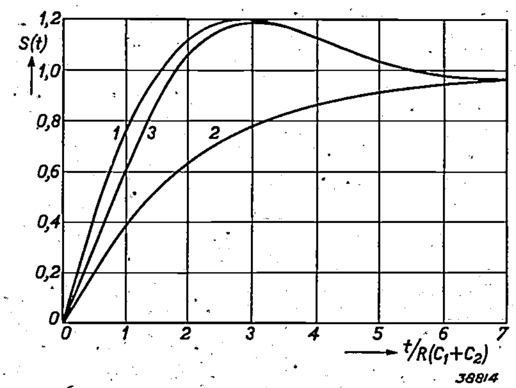


Fig. 10. Variation of the output amplitude in the case of single side-band amplification upon the use of a circuit with damping in parallel.
1 amplitude of the input current 0→1
2 amplitude of the input current 1→0
3 amplitude af the input current makes a small jump $i_3 \rightarrow i_3 + \Delta i_3$. In this case $S(t)$ stands for the output amplitude divided by Δi_3.

In *fig. 10* the behaviour of the amplitude modulation of the output signal as a function of the time for different input signals is shown for the case of single side-band reception. Curve *1* shows the behaviour with the transient 1→0. For the sake of comparison with the preceding curve, curve *2* is plotted in the reverse direction. It may be seen that there is here a very great difference in the shape of the curves, Curve *1* exhibits a peak for-

[6]) See the article referred to in footnote [4]).

mation of 20 per cent, while in the shape of curve *2* the oscillation is completely damped.

Curve *3* represents the variation of the amplitude modulation of the output signal when the jump in amplitude of the input signal is only small. In that case a jump in one direction and one in the opposite give symmetrical results, which approximately resemble curve *1*.

For television signals which are so modulated that the amplitude increases with increasing brightness of the picture (positive modulation) we may draw from curves *1* and *2* the conclusion that a transition from black to white will be sharper and will exhibit more peak formation than a transition from white to black. The latter has exactly the same behaviour in the output signal as if the carrier-wave lay in the middle of the sensitive band. Therefore if we assume that the worst transition determines the quality of the picture, we would reach the conclusion that the use of a single side band with a given total band width does not give better quality than the use of two side bands. If, however, we assume that the quality of the picture is determined chiefly by the reproduction of small contrasts (curve *3*), there is considerable improvement in the definition.

To what degree this may be considered as an improvement in the quality of the picture cannot be decided on the bases of our example. We considered an ordinary oscillation circuit which is analogous to a simple *R-C* circuit as far as its transient characteristic is concerned. Since in spite of this in the case of single side-band reception a 20 per cent overshoot occurs, it must be feared that a complete receiver will exhibit much greater oscillations, the combatting of which may be very difficult.

A mathematical treatment of single side-band reception with fairly complex networks is quite complicated, so that no detailed theoretical investigations have yet been carried out in this direction. It is, however, not dificult to determine the transient characteristics of receivers experimentally. Difficulties might here be expected because of the necessity of measuring a transient which occurs only once. This difficulty can, however, be avoided by applying to the input a block signal as in fig. 2 or a high-frequency signal with a block-shaped amplitude variation. If the period of this block signal is taken large enough the transition phenomena which occur due to a given transient will have become practically equal to zero by the beginning of the following transient. In this way the transient characteristic is found continually repeated, so that it can be made visible, for instance by means of a cathoderay oscillograph.

AN APPARATUS FOR TREATMENT WITH INFRARED RADIATION

by A. van WIJK.

621.384 : 615.83

Various local affections, such as rheumatism, can be successfully treated by heating the tissue by means of an irradiation apparatus. The wave length of the radiation can best be chosen so that the greatest possible amount of energy is absorbed with a given increase in temperature of the tissue close to the surface of the skin. This means that the absorption coefficient of the body must be low for the radiation in question. The most favourable radiation for this purpose is the region of red and infrared with wave lengths from 0.7 tot 1.3 μ. Radiation of longer wave lengths is highly absorbed especially by water, that of shorter wave lengths, by the red colouring matter in the blood.

In this article the Philips apparatus for infrared irradiation is described. The radiation is excited with the help of an incandescent filament lamp surrounded by a filter of running water. This apparatus furnishes practically exclusively radiation with the desired wave lengths, while in the case of the ordinary incandescent bodies of low temperature only a few per cent of the radiation has the desired wave lengths. On the basis of tests it is shown that with the new apparatus about three times as great an intensity of radiation can be applied as with the ordinary apparatus.

In the case of various local affffections (for instance rheumatism, inflammations) which do not lie too deep under the skin, but also in the case of diseases of the internal organs, use is made by doctors of the so-called infrared treatment, for curing or for relieving pain. It must be assumed that the favourable action is primarily a result of the heating of the tissue which occurs. Whether or not in addition a photochemical action of the radiation must also be assumed is uncertain, although this possibility is taken into account.

In the application of infrared irradiation the aim is as a rule to increase the radiation intensity as much as possible in order to heat the tissue in the the neighbourhood of the affection as highly as possible.

At all points in the tissue the heat freed per unit of time and per unit of thickness of layer is proportional to the local intensity of radiation. Since the radiation is strongly absorbed in the tissue and the intensity thus decreases rapidly with penetration, the heating effect in the layers first traversed is very much greater than at a somewhat greater depth. The natural limitation of the intensity of the irradiation applied is therefore furnished by the heating effect on the outermost layers of tissue, *i.e.* the skin and the layer directly beneath it, in which lie the organs of feeling. This heating may not exceed a definite value without causing intolerable pain.

The degree of heating depends not only upon the intensity but also upon the nature of the radiation employed, *i.e.* on the wave length, or the wave length region in which the energy is emitted. The wave length has a great influence on the absorption of the radiation in the tissue[1]). The slighter the absorption the greater the depth to which it acts and the greater the intensity of irradiation which can be applied before the absorption of energy in the outermost skin layer becomes so great that pain is felt (*cf. fig. 1*). It is therefore important to choose a type of radiation for which the absorption is as small as possible.

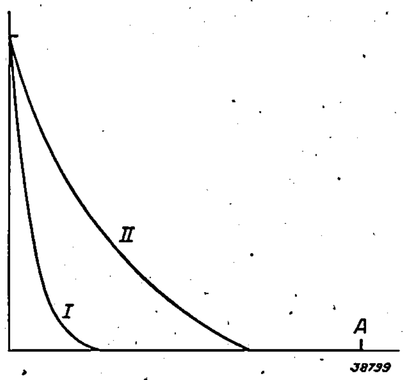


Fig. 1. Diagram of the variation of the energy absorbed per unit of time and per unit of thickness of layer as a function of the depth in the tissue,
I for radiation with high absorption,
II for radiation with slight absorption.
The irradiation intensities are so chosen that the absorptions directly under the skin are equal. The surface of the curves gives the total energy absorbed per unit of time, which therefore in the case of curve *II* is considerably greater than in the case of curve *I*. For a point deeper than the absorption region, *A* for instance, this energy determines the heating effect which takes place by conduction.

In practice this requirement is not at all satisfied at present in infrared treatment. The most commonly used irradiation apparatus contains as source of radiation an electrically heated filament, in the open air, whose temperature varies from 500 °C (barely incandescent) to 1 000 °C (orange-red incan-

[1]) *Cf.* for example G. Miescher, Strahlenther. **61**, 578, 1938.

descence). The relative composition of the radiation of such bodies is approximately described by Planck's law for the radiation of "black bodies" [2]).

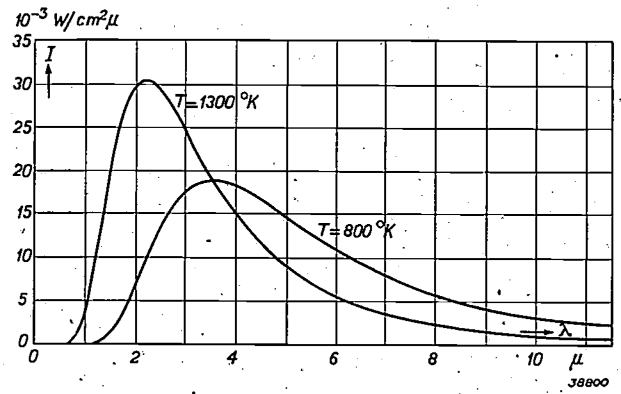


Fig. 2. Irradiation intensity in W/cm² per wave-length region of 1 μ as a function of the wave length, for black bodies with temperatures of 800 °K and 1 300 °K, with a total irradiation intensity of 0.1 W/cm².

Fig. 2 gives the variation of the radiation intensity as a function of the wave length for the temperatures mentioned, which correspond to absolute temperatures of 800 °K and 1 300 °K [3]). The curves are drawn for cases of equal irradiation intensity, namely 0.1 W/cm². The maxima of the radiation intensity lie far in the infrared for both temperatures, at wave lengths (λ) of about 3.6 and 2.2 μ, respectively. Only a negligibly small fraction of the radiation lies in the visible region ($\lambda<0.8$ μ). This is in agreement with the customary name of "infrared treatment" as compared with the "heating by means of radiation" considered in this article.

In principle, however, radiation of other wave-length regions might just as well be used, since the heating effect of the absorbed radiation is the same for all wave lengths, namely 860 kcal/kwh. The fact that infrared radiation has been chosen, and that sources of radiation of low temperature have thus been arrived at for generating this radiation is based upon the very wide-spead misconception that only infrared radiation furnishes heat. which misconception is expressed in the term "heat radiation" for infrared. The origin of this misconception may be sought in the fact that in the case of practically all existing sources of radiation (electric lamp; candle, fire) by far the greatest part of the energy emitted lies in the infrared spectral region, so that what one feels of the radiation is mainly due to this region. This is, however, only a question of quantity and not of quality.

[2]) For practical applications of Planck's radiation formula see for example the tables of W. de Groot, Physica **11**, 265, 1931 or Jahnke-Emde, Funktionentafeln, Teubner (Leipzig, Berlin) 2nd edition, 1933 p. 46.

[3]) Absolute temperature in °K = temperature in °C + 273°.

The lower the temperature the greater the proportion of the radiation emitted in the infrared. On the basis of the misconception indicated, and in the mistaken expectation of a high efficiency, this has led to the use of sources of radiation of low temperature.

The problem is not, however, to obtain the highest possible percentage of infrared radiation, but to choose a kind of radiation which permits the highest possible load on the skin, *i.e.* as already explained, a kind of radiation which is absorbed as little as possible by the tissue. If we now recall that the living tissue consists for the most part of water, we see that the radiation which is strongly absorbed by water cannot have a very great depth of penetration into the tissue. *Fig. 3* shows the absorption curve [4]) of water, as well as the transmission calculated from it with thicknesses of layers of 1, 5 and 10 mm. It is clear that even with thin layers practically no radiation of $\lambda>1.35$ μ is transmitted

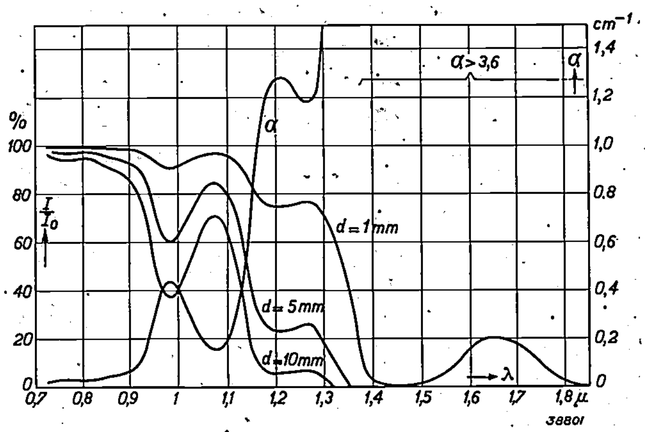


Fig. 3. Absorption index a and transmission I/I_0 for water at different thickness of layer d. The absorption index a is defined by the relation:

$$\frac{I}{I_0} = 10^{-ad}.$$

At wave lengths greater than 1.35 μ water exhibits strong absorption.

(with a thickness of 1 mm a transmission region with a maximum of 20 per cent may be observed at 1.65 μ, at 2 mm the height of this maximum has been reduced to 4 per cent, at 3 mm to 0.8 percent). It may therefore be concluded immediately that radiation with $\lambda>1.35$ μ can never reach the deeper layers of tissue, but is entirely absorbed in the outer layer, and that it is therefore unsuitable for the therapy in question.

Radiation of too short wave lengths is also unfavourable for the therapy in question. For example, ultra violet light is strongly absorbed and the same

[4]) J. R. Collins, Phys. Rev. **26**, 771, 1925.

is true of visible light with the exception of the extreme red. The colouring matter of the blood is to a large extent responsible for this absorption. Direct transmission tests with layers of tissue have shown that the maximum of transmission lies between 0.7 and 1.35 μ [5]), whereby the long wave end of the curve corresponds to that of water, and the short wave end to that of oxyhaemoglobine [5]), *i.e.* the red blood colouring matter in the oxidized form (arterial blood).

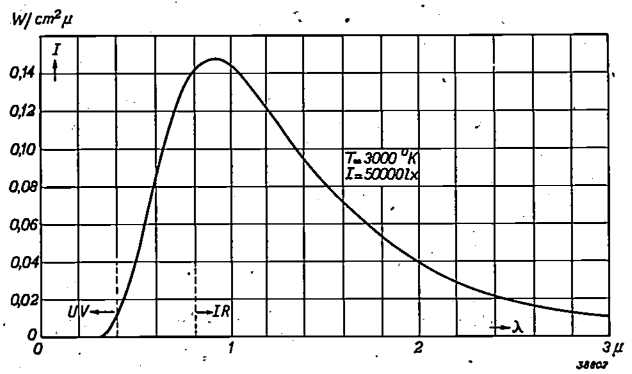


Fig. 4. Irradiation intensity as a function of the wave length for an incandescent body of tungsten with a temperature of 3 000 °K, with an illumination intensity of 50 000 lux.

The radiation emitted by black bodies at low temperature is very unfavourable in connection with the transmission. At 1300 °K only 4 per cent of the radiation has a wave length shorter than 1.35 μ; at 800 °K this proportion is even less than one per thousand (*cf.* fig. 2). By choosing a higher temperature, however, it is found possible to obtain radiation of which a much larger percentage belongs to the wave-length region of maximum transmission. *Fig. 4* shows the distribution of the radiation from tungsten at a temperature of 3 000 °K. The curve is valid for a case in which the illumination intensity amounts to 50 000 lux. In this case the radiation density is 0.168 W/cm² (calculated to $\lambda = 3$ μ; the tungsten filament is always surrounded by a bulb of glass or quartz which absorbs radiation of longer wave length). The distribution of the radiation over the different wave-length regions is shown in the following table.

Table I

wave-length region	radiation density	
	W/cm²	%
infrared $\lambda > 1.35$ μ	0.062	37
infrared 1.35 μ $> \lambda >$ 0.7 μ	0.088	52
visible 0.7 μ $> \lambda >$ 0.4 μ	0.018	11
ultraviolet $\lambda < 0.4$ μ	$0.45 \cdot 10^{-5}$	0.27

As the table shows, 0.088 W/cm² of the radiation, *i.e.* 52 per cent, falls within the favourable region from 0.7 tot 1.35 μ, in addition, however, 37 per cent falls in the unfavourable region $\lambda > 1.35$ μ which is too much absorbed on the surface of the tissue. By filtering the radiation through a layer of water, that part which contains the unfavourable wave lengths can be removed in advance, and the composition becomes much better, in the sense that a still higher percentage lies in the region 0.7—1.35 μ. *Fig. 5* shows the spectral distribution of the radiation behind water filters of 5 and 10 mm, respectively. The percentages in the region 0.7-1.35 μ are 74 per cent and 70 per cent, respectively; practically all of the remainder lies in the visible. As may be seen from the decrease in the percentage upon increase in the thickness of the water layer from 5 to 10 mm, a further increase in this thickness is of no advantage.

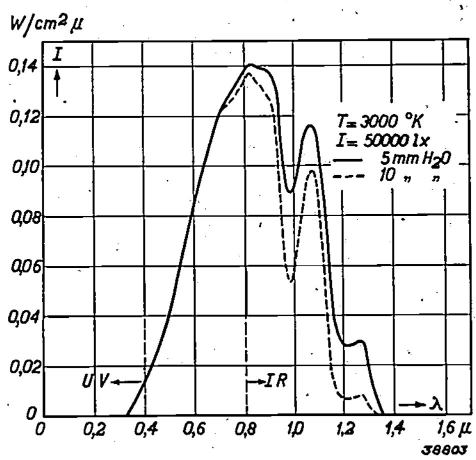


Fig. 5. Irradiation intensity as a function of the wave length for a tungsten wire at 3 000 °K with water filter, at an illumination intensity of 50 000 lux. Continuous line, water filter 5 mm thick. Broken line, water filter 10 mm thick.

The Philips apparatus for infrared irradiation

The Philips apparatus for infrared irradiation (*cf. fig. 6*) contains an electric lamp of 750 W with a filament temperature of 3 000 °K, surrounded by a layer of water about 6 mm thick. It is necessary to use running water, since otherwise it would quickly boil, due to the large quantity of energy it absorbs. A comparison of the curves of figs. 4 and 5 shows that with a layer of water 5 mm thick about 54 per cent of the energy emitted is absorbed in the water, in addition to which is the heat which reaches the glass bulb of the lamp by convection and the energy absorbed in the bulb, both of which can also be given off to the water. Direct measurement by determination of the temperature differ-

[5]) *Cf.* G. Hoffmann, Strahlenther. **65**, 477, 1939; also U. Henschke, Strahlenther. **66**, 646, 1939; this survey, from which various data have been borrowed, contains a detailed bibliography.

ence of the incoming and outgoing water gave an energy of 560 watts absorbed by the water.

By means of an adjustable reflector the radiation can be more or less concentrated as required. For

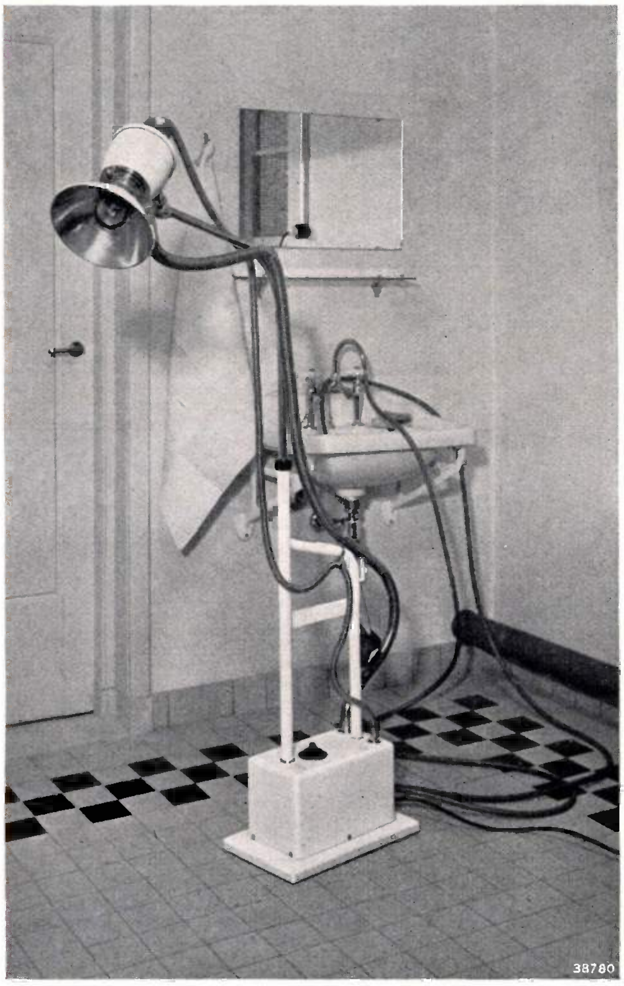


Fig. 6. Philips apparatus for infrared treatment.

the sake of this possibility it is also important that the temperature of the radiation element should be high, so that its dimensions need only be small. In our case the largest dimension is about 1 cm. Since the total radiation flux emitted per unit surface of the source is proportional to the fourth power of the temperature, at a temperature of 800 °K the surface area would have to be $(3\,000/800)^4$ = about 200 times as large for an emission of the same energy, and the reflector would also have to be correspondingly very much larger in order to obtain the same concentration of beam and adjustability of beam formation.

The lamp is fastened to an adjustable standard, similar to that of the Philips "Biosol" apparatus [6]). The lamp (with reflector) can be rotated through angles of 60° about two mutually perpendicular axes, so that the beam can easily be directed upon the desired part of the body. In the case of the standard there is a transformer with separated windings, by which the mains voltage (A.C. voltage 220 V) is transformed to 15 volts. The lamp current (50 A) is conducted to the lamp through a cable surrounded by a flexible metal sheath; through a second thick metal tube run the rubber inlet and outlet tubes for the filter water.

Properties of different sources of radiation

It was noticed by Henschke [7]) that the radiation tolerance of the skin (*i.e.* according to the above the maximum intensity of the incident radiation which can be borne without pain), in contrast to many other subjective quantities, can be fairly sharptly determined, and with a given composition of radiation it varies very little for different individuals. During short times very high irradiation intensities can be borne, but as the time of irradiation increases the permissible intensity decreases and approaches an asymtotic value. In *fig.* 7 the experimentally found time during which a given radiation intensity can be tolerated is plotted as a function of this intensity for radiations of different composition [8]). It is evident how relatively sharply

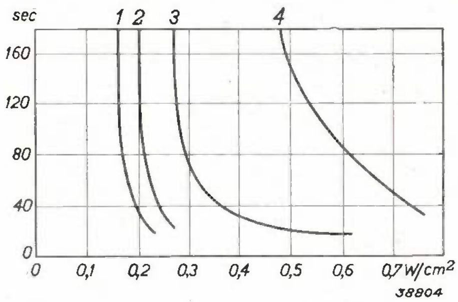


Fig. 7. Time during which the skin can tolerate a radiation as a function of the irradiation intensity, for the four different sources of radiation of table II.

the limit is defined of the radiation intensity which can be tolerated for a long time (*i.e.* longer than 3 min.) and how widely the skin tolerance for radiation differs for the different kinds of radiation.

In *table II* the tolerances deduced from fig. 7 are given for long irradiation times.

According to the table, with the Philips apparatus for infrared irradiation a skin dosage can be given which is approximately 3 times as great as the

[6]) See Philips techn. Rev. 2, 18, 1937.

[7]) See the article referred to in footnote [5]).

[8]) The values found by us deviate appreciably from the figures given by Henschke (footnote [5]). The latter are about 30 times as large. It seems probable to us that the values given by Henschke are given per minute and not per second as he states.

infrared sources now being used, and consequently a correspondingly greater heat effect can be attained. The nature of the radiation is such that the maximum depth effect is obtained. Nevertheless in this case also at any great depth (several mm or more) the heating effect is mainly due to conduction and not to radiation which has penetrated that far, since the latter is already much too attenuated (see fig. 1).

Table II

Source of radiation	Temperature.	Skin tolerance for long-continued irradiation
1) Ordinary infrared irradiation apparatus (radiator)	1 300 °K	0.16 W/cm²
2) Electric lamp under low load	2 200 °K	0.20 W/cm²
3) Electric lamp under high load	2 900 °K	0.27 W/cm²
4) Philips' infrared irradiation apparatus	3 000 °K	0.47 W/cm²

The temperature increase as a function of the depth in the tissue has been determined by different investigators. It may be seen from the measurements that at a greater depth than about 10 mm no increase in temperature can be observed, even with the optimum composition of the radiation. Since infrared irradiation is also often used for more deeply lying affections, it must be assumed that if there is any effect at all it cannot be caused by direct heating, but by an indirect effect, for instance reflective. Another possibility of explanation is that there is photochemical action and to an important extent. Even at a great depth where the attenuation is so great that there can be no question of a heat effect when radiation in the region 0.7-1.35 μ is considered, the intensity is still great enough to exert a possible photochemical effect.

It cannot reasonably be expected that radiation with a wave length longer than 3 μ can be important in such a photochemical effect; the radiation quanta in that region are so small that they can only promote chemical reactions which require such a low activation energy that they would also be brought about by thermal agitation. Therefore its photochemical action plays any essential part in infrared irradiation, the radiation which is furnished by the Philips apparatus will be favourable, not only because of its greater depth of penetration, but also because of the greater energy of the radiation quanta.

It is interesting to note that the radiation of the sun is also filtered by water before it reaches the earth, and to about the same degree as that chosen for the Philips aparatus. The amount of water vapour in the earth's atmosphere is such that if it were all condensed a layer would be obtained whose thickness would be of the order of 1 cm. Thus the amount of water in the atmosphere above the Astrophysical Observatory on Mt. Wilson corresponds to an average thickness of 0.69 cm, while the extremes are 0.2 and 2.8 cm. At the long wave end the boundary of the spectrum of the Philips apparatus for infrared irradiation is about the same as that of the sun's spectrum. The maximum of the radiation intensity of the sun lies in the visible region (about 0.5 μ), that of the Philips apparatus at about 0.8 μ.

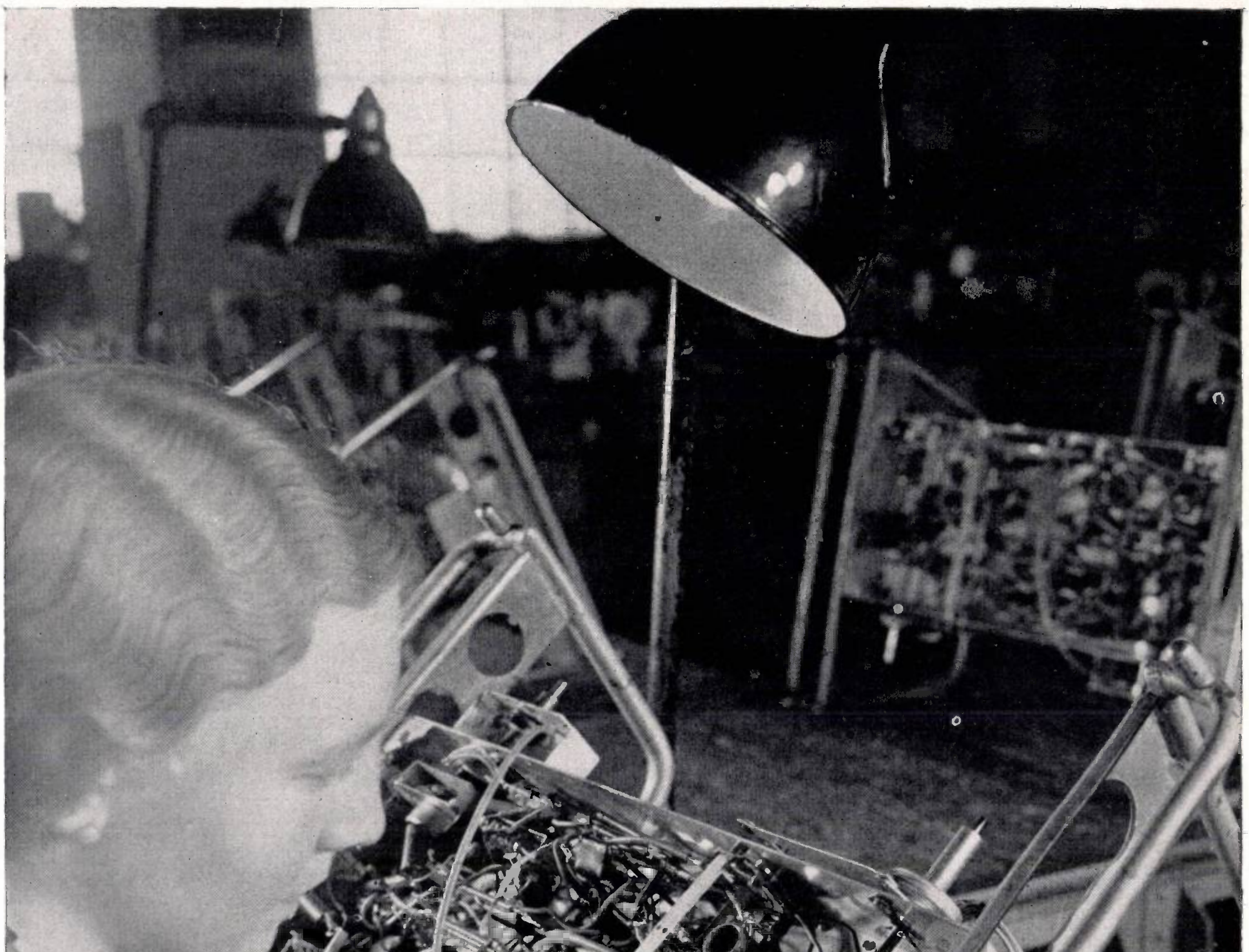
About 35 per cent of the sun's radiation lies in the wave-length region of greatest transmission (0.7-1.35 μ), with the Philips apparatus this figure is about 70 per cent, so that the tolerance of the skin for the sun's radiation will be lower than for that of the apparatus.

The intensity of the sun's radiation on earth amounts to about 0.1 W/cm² under favourable conditions, *i.e.* it is of the same order of magnitude as was found in table II for the skin tolerance with different compositions of radiation. If the intensity of the sun's radiation is increased to for instance two or three times the original value by means of a concave mirror, the limit of the skin tolerance is actually reached.

The cooling due to air currents is of great importance on the experimentally found tolerance. Even with quite a low wind strength more can be tolerated than in a perfect calm. The fact that on certain days the sun can "burn" so much is probably rather a result of calm air than of a particular spectral composition of the sunlight (due for example to an extremely small water content of the atmosphere).

In the application of artificial radiation also, by cooling the skin either by a current of air or by a glass cuvette with water cooling pressed against the skin (compressor), the radiation tolerance can be considerably increased. This fact is used by some doctors; due to the great influence of the heat of conduction on the temperature distribution resulting in the tissue, however, this does not immediately bring about an increase in the temperature obtained at greater depths. Upon strong cooling (compressor with water) of the surface of the skin, a decrease may even occur. Any possible photochemical effect, however, can in this way be increased in any case, since the permissible irradiation intensity becomes greater.

TESTING THE CONTACTS IN A RADIO RECEIVER



The large number of connections in a radio receiver made by soldering, welding or in any other way, where a fault in one single connection may upset the functioning of the whole set, makes various kinds of tests essential. Moreover, care must be taken that the manner of making the connections ensures good quality for a long time. To fulfil this requirement a large number of checks and random tests are unavoidable, however time-consuming they may sometimes be.

A DISCHARGE PHENOMENON IN LARGE TRANSMITTER VALVES

by J. P. HEYBOER.

537.521.7 : 621.396.615.1

In transmitter valves a flashover sometimes occurs between the electrodes under high voltages in spite of a good vacuum (Rocky Point effect). In large valves this phenomenon may result in considerable damage, especially to the filaments, if no precautions are taken. The damage can be explained qualitatively, and to some extent also quantitatively by the ponderomotive forces which occur between different branches of the filament as a result of the current which flows through these branches when such a flashover occurs. The article gives several simple calculations connected with this and describes several experiments with a model in which the phenomena observed in practice could be imitated. In conclusion the way in which the harmful results of a flashover can be avoided are discussed.

In modern transmitter valves there is a vacuum of 10^{-7} to 10^{-8} mm Hg. Not only is the high vacuum of importance for the satisfactory functioning of the cathode, which must furnish the emission current in the valve and upon which gases would exert a harmful effect, but also for obtaining adequate insulation between the different electrodes. This is of particular importance since in the largest water-cooled transmitter valves peak voltages of about 40 kV may occur, while the distance between the electrodes, for example between anode and grids, amounts to only a few centimetres. With too high a gas pressure the ionization of the gas molecules by the fast electrons would quickly lead to the occurrence of a continuous gas discharge in the valve. The vacuum is most simply controlled by measuring directly the ion current which flows to the control grid (which is at a negative voltage) as a result of the ionization.

From time to time the insulation between the anode and the other electrodes suddenly disappears and a flashover occurs inside the valve. This occurs in such a way that neither before nor after such a flashover is any increased ion current to the grid measured. The phenomenon may not therefore be ascribed to a gradual depreciation of the vacuum, but it must be assumed that during the functioning of the valve, due to some cause or other, a small quantity of gas is suddenly freed from one of the electrodes. This results in a flashover, while directly afterwards the gas disappears again, probably by absorption in the electrodes or in the getter which is usually present.

The phenomenon described has long been known and is usually indicated in the literature under the name of "Rocky Point effect" from the name of an American transmitting station where the phenomenon was first observed.

Wat are the results of the Rocky Point effect on the transmitter valve? Low power valves, for instance of 100 W, withstand such a brief discharge without the slightest trouble. In the case of valves of high power, however, the discharge may lead to damage. This is understandable when the connections and construction of these valves are considered.

The anode voltage of a transmitter valve is usually supplied by a rectifier with a smoothing filter, which filter is terminated by a condenser of fairly large capacity (in large valves 20-40 μF). Between the anode and the rectifier a choking coil is further included in order to prevent the high-frequency voltage of the anode from reaching the rectifier. The supply arrangement thus takes on the form shown in *fig. 1*. In ordinary use the condenser C is charged to the anode D.C. voltage, which in large valves may amount to 10 to 20 kV.

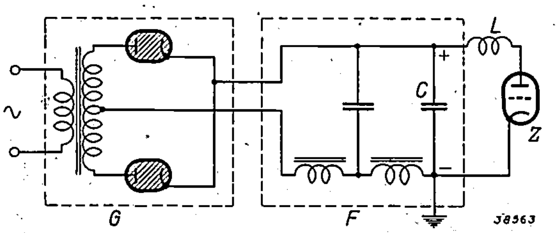


Fig. 1. Diagram of the supply arrangement of a transmitter valve (Z). G rectifier, F smoothing filter.

At the moment when a flashover occurs in the valve, the anode-cathode space functions as a short circuit. The condenser is thus discharged *via* the choke L and very large discharge currents may occur which seek a path to earth through the valve. A consideration of *fig. 2a* and *b* in which the construction of a large water-cooled transmitter valve, and particularly the construction of the cathode, is shown diagrammatically furnishes some insight into the path which the discharge currents will choose. The cathode is a tungsten filament which consists as a rule of two or more branches connected in parallel. The wires of each branch are strung in a zigzag in the direction of length of the valve, while about them, arranged around the circumference of a cylinder, wires of one or more cylindrical grids and finally the cylindrical anode are placed. In fig. 2*b*, in which the cathode is shown flattened out, it may be seen that the beginnings of both branches come together in an earthed supply terminal P, and in the same way the ends in a second supply terminal. Upon a flashover from anode to earth the current will probably now go mainly to the wires AP and BP and flow off to earth through these wires.

Upon the occurrence of a flashover therefore the

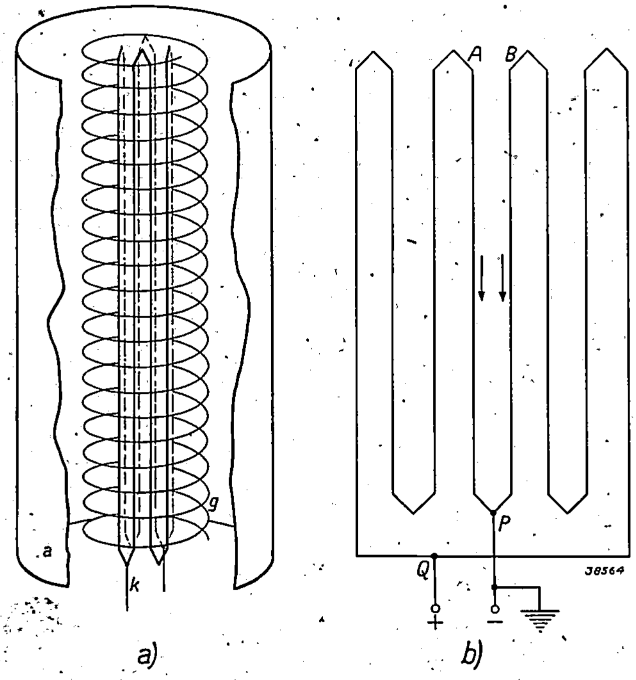


Fig. 2. *a*) Electrode system of a large water-cooled transmitter valve (triode). Around the filaments k which are strung as lines on the development of a cylinder, the control grid g is wound in the form of a spiral, while about that in turn is placed the cylindrical anode a.

b) Flattened out cathode of the water-cooled transmitter valve in question; it consists of two branches connected in parallel, each having four tungsten filaments. The two branches are fed by the common terminals P and Q, P being earthed.

adjacent wires AP and $B\ P$ may conduct very high currents; as a result there is not only considerable heat development, but at the same time large ponderomotive forces occur between the two wires. It is easily understandable that these forces may cause appreciable deviations or even permanent changes in shape of the wires.

In practice indeed such a deformation of the filament has several times been observed after a flashover, and the deviation of the wires was sometimes such that mutual contact or contact with the control grid occurred. In one case even one of the wires was torn free at the terminal. At the same time traces of the discharge were observed on the filaments, sometimes in the form of many-branched figures, usually as irregular cavities in the wires.

While these facts are explained qualitatively by the above sketched mechanism, it is also desirable to obtain a more quantitative picture of the phenomenon. The calculations and experiments undertaken to this end will be discussed briefly in the following, while a discussion will also be given of the way in which the damage can be avoided.

The whole calculation may be divided into three parts: 1) the discharge currents; 2) the forces hereby exerted on the filaments, 3) the deviations of the wires thereby caused. (The effect of the heat development will not be considered).

Calculation of the discharge current

For the calculation of the discharge current we begin with the diagram of fig. 1 in which the condenser C is charged to a voltage V and at the moment $t = 0$ is dicharged *via* the self-induction L and a resistance R. The latter, in ordinary connections, consists only of the loss resistances in coil and condenser and of the resistance of the filaments AP and BP in parallel. We assume that these filaments are traversed by the current in their full length, from A to P and from B to P.

With the help of alternating current theory the following expression is found for the discharge current $i(t)$:

$$i = i_m \sin \beta t\, e^{-at}, \qquad (1)$$

where

$$\left.\begin{aligned} i_m &= V/L\beta, \\ a &= R/2L, \\ \beta &= \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}. \end{aligned}\right\} \qquad (2)$$

According as a real or an imaginary value is obtained for β (the latter is the case according to equation (2) only with sufficiently large damping R), the expression (1) will represent a damped sinusoidal vibration or an aperiodically varying function.

As an example let us consider a case in which defects were actually caused by the Rocky Point effect. In this case $V = 12$ kV; $L = 1\,950$ μH; $C = 32$ μF. Each filament (at the working temperature of $2\,500$ °K) had a length $l = 10.63$ cm and a diameter $d = 0.0892$ cm; with the specific resistance of tungsten at the working temperature (73.9×10^{-6} ohm cm) a value of about 0.07 ohm is calculated for the resistance of the two wires in parallel. As the total resistance of the dicharge circuit (in which are included the loss resistances C and L) we therefore assume the round value $R = 0.1$ ohm. With these values we obtain

$$i_m = 1\,540 \text{ A}; \alpha = 25{,}6 \text{ sec}^{-1}; \beta = 4\,000 \text{ sec}^{-1}.$$

The current is thus actually found to reach very high values. Its variation with time (equation (1)) is shown in *fig. 3*. It may be seen that the discharge has a rapidly oscillating character (frequency $\beta/2\pi = 640$ c/s), while the damping is such that the current amplitude after about 0.3 sec., *i.e.* 20 cycles of the oscillation has fallen to one half.

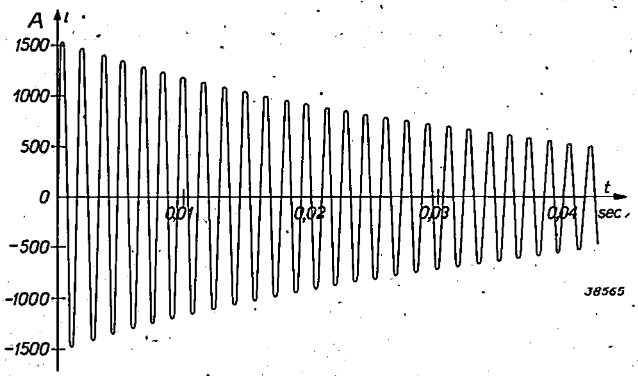


Fig. 3. Variation of the current i upon discharge of a condenser of 32 μF charged to 12 kV, *via* a self-induction of 1 950 μH and a resistance of 0.1 ohm.

Calculation of the mechanical forces

We shall assume that the current is divided equally between the two wires, so that each wire bears the current $i/2$. If r is the distance between the wires the magnetic field strength which the current in one wire exerts on the other wire is

$$H = \frac{i}{r}.$$

Due to the fact that this other wire itself carries the current $i/2$, it experiences a force per unit length of

$$p = H \cdot \frac{i}{2} = \frac{i^2}{2r} \text{ dynes/cm},$$

when r is given in centimetres and i in electromagnetic units. If we measure i in amperes, then

$$p = \frac{i^2}{200\,r} \text{ dynes/cm}. \quad (3)$$

Since the current has the same direction in both wires, the direction of p is such that the wires are always drawn toward each other. If in (3) we fill in the expression (1) for the current, we obtain

$$p = \frac{i_m^2}{400\,r} (1 - \cos 2\beta t)\, e^{-2\alpha t} \quad (4)$$

In *fig. 4* this variation of the force with time is plotted with $r = 0.617$ cm as in the example considered. If we neglect the damping for the time being the force may be described roughly as follows:

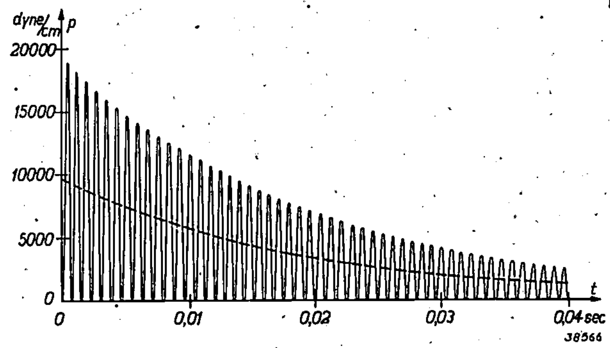


Fig. 4. Variation of the force p which is exerted by the discharging current of fig. 3 on 1 cm of each of the two filaments AP and BP.

from the moment $t = 0$ a constant force of the magnitude

$$p_0 = \frac{i_m^2}{400\,r}, \quad (5)$$

acts upon each of the two wires, while upon this is superposed a force with the amplitude (5) varying at the frrquency $2\beta/2\pi = 1\,280$ c/s. With the given values of iF and r, p_0 becomes 9 620 dynes or about 10 g per cm length of the wire.

The damping is here twice as great as for the current, so that the force has fallen to one half after about 0.15 sec.

The bending of the filaments

The filaments may be considered as rods which will be set vibrating by the force (4). The quantity in which we are interested is the maximum deviation which the rod thereby makes. The fact that these deviations will not be only very small can easily be understood when it is considered that the wire in our case, according to the data in *table I* weighs about $1^1/_4$ g, while at the beginning an

Table I

Data about the filaments of the valve in question for the working temperature of 2 500 °K.

Length	$l = 10.63$ cm
Thickness	$d = 0.0892$ cm
Cross section	$q = \pi d^2/4 = 0.00622$ cm^2
Density	$\varrho = 18.6$ g/cm^3
Mass	$m = \pi d^2 l \varrho/4 = 1.24$ g
Modulus of elasticity	$E = 2.83 \cdot 10^{12}$ dynes/cm^2
Equatorial moment of inertia	$I = \pi d^4/64 = 3.11 \cdot 10^{-6}$ cm^4
Separation of the wires	$r = 0.617$ cm

average load acts upon it of $l{\cdot}p_0$ = about 100 g, *i.e.* 80 times its own weight.

In order to obtain an idea about the bending of the wire in a simple way let us consider it as a rod, which is supported at both ends. Furthermore we assume that the rod has the same elastic properties throughout its whole length [1]). For such an elastic system the frequencies at which it will come into resonance can be calculated according to the formula

$$f_n = \frac{(2n+1)^2\pi}{2\ l^2}\sqrt{\frac{EI}{\varrho q}}, \text{ with } n = 0,1,2, \ldots \quad (6)$$

The length of the rod l occurring herein, the modulus of elasticity E, the equatorial moment of inertia I of the cross section of the wire, the density ϱ and the cross section q (all at the working temperature) can be taken from table I. One then finds:

$$f_0 = 121 \text{ c/s},$$
$$f_1 = 1\,085 \text{ c/s},$$
$$f_2 = 3\,020 \text{ c/s, etc.}$$

None of these frequencies lies in the immediate neighbourhood of the frequency of the varying force component (1 280 c/s). Resonance phenomena need not therefore be expected, and the order of magnitude of the result will not be affected if for the sake of simplification we entirely neglect the varying term of the force (4). Furthermore it may be seen that the rod vibrating at its fundamental frequency f_0 has already completed almost two full cycles before the force has fallen to one half. In calculating the maximum deviation therefore no large error will be made if the picture is further simplified by also neglecting the damping of the force [2]).

The problem is now reduced to the case represented in *fig. 5*: a uniformly distributed load of the magnitude p_0 (equation (5)) is applied at the moment $t = 0$ to a homogeneous rod supported at both ends.

Such a suddenly applied load always causes a greater deviation than the same load in the equilibrium condition (static load). In oscillation systems with a single degree of freedom, for example, it is known that this makes a difference of a factor 2. We shall therefore find a lower limit for the expected deviation of the rod if we simply as-

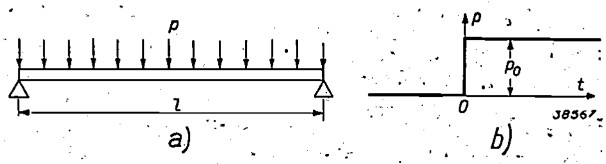


Fig. 5. The filament is considered as a homogeneous rod supported at the ends, on which acts a uniformly distributed load p (a) of the character shown in (b).

ume a static load of the magnitude p_0. The following formula is then valid for the greatest deviation δ occurring at the middle of the rod:

$$\delta = \frac{5}{384}\cdot\frac{p_0 l^4}{EI} \quad (7)$$

With the data of table I it follows from this that

$$\underline{\delta = 0.183 \text{ cm}} \quad (8)$$

If we now calculate the maximum tensile stress δ occurring at the middle of the rod, for which in the case of fig. 5a the following formula is valid:

$$\sigma = 4.8\,\frac{Ed}{l^2}\,\delta \text{ dynes/cm}^2,$$

we find a value of about 2 000 kg/cm². Since the tensile strength of tungsten at the working temperature of 2 500 °K is only 470 kg/cm² [3]) the material will begin to yield long before the deviation given by (8) has been reached. The wires will therefore either break or undergo permanent deformations which may be considerably greater than the bend according to (8). Since already at $\delta = (r-d)/2 =$ 0.265 cm there is contact between the two wires which attract each other, it is of little use to give a more exact calculation for the rough estimation here given [4]). For a slightly different situation, however, a more exact calculation will be given below.

---

[1]) The first assumption is not entirely correct, since the filament is clamped at one end, while at the other end the fastening is intermediate between clamping and supporting. At the latter point the wire passes by means of a kink over into the adjacent wire. This kink is kept in place by a hook. The second assumption, that of homogeneity of the rod, is also not entirely correct, since the temperature of the filament is somewhat lower at its ends due to heat conduction from the supply terminal and hook, so that the modulus of elasticity is higher than in the middle.

[2]) A more exact calculation in which the damping was not neglected gave as a result that the maximum deviation already occurs after 0.004 sec, *i.e.* after even less than half a period of the fundamental vibration.

[3]) According to W. Espe and M. Knoll, Werkstoffkunde der Hochvakuumtechnik, J. Springer, Berlin 1936, fig. 15.

[4]) It must be pointed out that the various simplifications which are introduced have mutually opposite effects: the damping of the force, the fact that the current may only flow through a part of the wire, that the wire is more or less clamped and is somewhat stiffer at its ends due to the lower temperature, all these facts make the deviation smaller. That the loading is not static, but suddenly applied, makes the deviation greater, in the same way the varying force component contributes to this. All in all it is fairly certain that the region of yield will be reached, and this is already sufficient for the train of thought here followed.

Experiments

While the calculation given above already makes it clear that the damage observed due to the Rocky Point effect can indeed be explained by the simple mechanism proposed, it nevertheless still seemed desirable to obtain a graphic confirmation of this by imitating the effect experimentally. For this purpose the same type of filament as in the example considered was mounted in a glass bulb, while with the connections represented in *fig. 6* a condenser of 540 μF which was changed to 1 000 volts could be discharged via the two parallel branches of the filament. Here also a clearly visible permanent bend in the wires occurred upon discharge. By repeating the discharging of the condenser several times the bend could be made greater and greater, as may be seen in the photographs of *fig. 7a-d.* If afterwards the direction of the discharge current in one of the two branches of the filament was reversed (the two branches had separate leads out of the bulb for this purpose, see *fig. 8*), the wires which then repel each other could be forced apart again to the original condition by a number of discharge impulses.

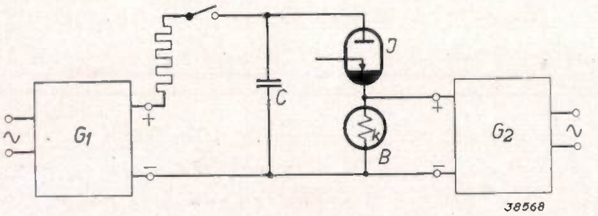


Fig. 6. Connections for the experiments with a model. The filaments k are placed in a bulb B. The condenser C is charged to 1 000 volts by the rectifier G_1, and can be discharged *via* the relay valve J and the filaments k. The relay valve, a glass vessel with a carbon anode, an auxiliary electrode and a pool of mercury as cathode, flashes over when an impulse voltage is applied to the auxiliary electrode; it serves here as a switch for the very high currents. By means of the rectifier G_2 the filaments are heated in order better to imitate the working conditions (smaller modulus of elasticity, lower yield value).

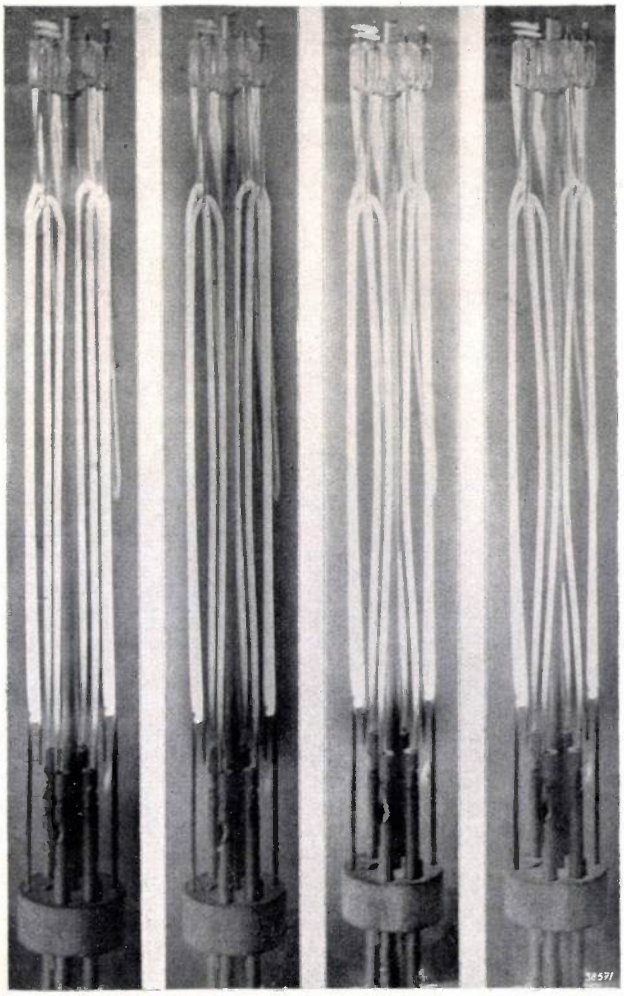


Fig. 7. After a number of discharge impulses a clearly visible permanent deformation of the filaments had occurred: *a*) initial condition, *b*) after 6 discharges, *c* after 12 discharges, *d*) after 18 discharges.

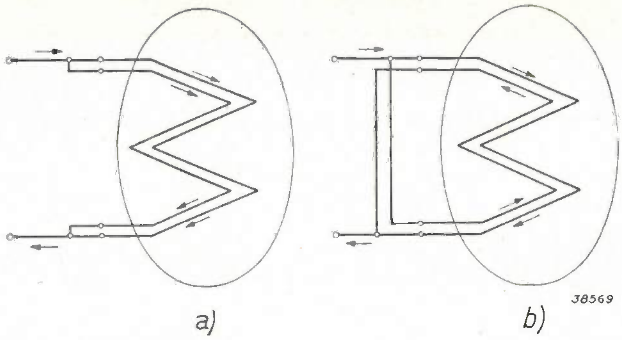


Fig. 8. The two branches of the cathode are led out of the bulb separately, in order to be able to send the discharge currents through the adjacent wires in the same direction (*a*) as well as in opposite directions (*b*).

In these experiments, where the mechanism which was assumed as an explanatory one is as it were isolated, and other phenomena which might play a part in the complex structure of a transmitter valve are excluded, the typical deformations of the Rocky Point effect actually do occur.

Combatting the damage

The obvious question now is what must be done to avoid the defects described? As long as it is impossible to attack the evil at its root (the spontaneous freeing of a small amount of gas) attempts must be made at least to eliminate the harmful effects of the discharge. This can be done very simply by including in the supply line of the anode a resistance of adequate size. This resistance is connected directly in series with the choking coil L (see fig. 1).

If in this way the total resistance of the discharge circuit is made 40 ohms, for instance, then in formula (2) the "angular frequency "β becomes imaginary, namely $\beta = j \cdot 3\,200$. The discharge is therefore no longer oscillating but aperiodic. Equation (1) now becomes

$$i = i_m' \,[e^{-(\alpha-\gamma)t} - e^{-(\alpha+\gamma)t}], \quad . \; . \quad (9)$$

with $\gamma = \beta/j = 3\,200\,\mathrm{sec}^{-1}$, $i_m' = \dfrac{V}{2\,L\gamma} = 960$ A,

$$\alpha = \frac{R}{2\,L} = 5\,100\,\mathrm{sec}^{-1}.$$

In *fig. 9* this curve is plotted. The peak value of the current, which is reached after about 1/4 000 sec, is in this case also still considerable, namely about 480 A; there is, however, a very much greater damping than in the above discussed case where $R = 0.1$ ohm: the current has practically disappeared after 0.002 sec. For the force which acts per cm length of the filaments we find with (3) and (9):

$$p = \frac{i_m'^2}{200\,r}\,[e^{-2(\alpha-\gamma)t} + e^{-2(\alpha+\gamma)t} - 2e^{-2\alpha t}]. \quad (10)$$

This is a similar curve to that in fig. 9 with a still greater damping. Thus in this case a force is exerted on the eleastic rods for only a very short time, and consequently only relatively small deviations will occur.

Since the loading now no longer resembles the static case in the least, the calculation of the deviations should be carried out by means of the differential equation of the vibrating rod. This calculation is sketched briefly at the end of the article. We may however, also obtain a good idea of the expected deviation again by a rough estimation.

Due to the fact that the force acts for such a short time we may assume that the whole impulse $P = \int pl\,dt$ is applied to the rod while it is still at rest. All the particles of the rod then take on a certain initial velocity v and the rod therefore receives a kinetic energy $K = \int {}^1\!/_2\, v^2\,dm$. If we assume that all the particles of the rod move in phase (fundamental vibration of the rod), the rod will have reached its greatest deviation only after a quarter period of the vibration, the kinetic energy is at that moment entirely converted into work of deformation U. If the shape of the rod in vibration corresponds to that upon static bending — an assumption which will certainly not be far from the truth —, there is the following relation between the work of deformation U and the deviation δ at the middle:

$$\delta = \frac{5}{384}\sqrt{\frac{240\,l^3}{E\,I}\,U} \quad . \; . \; . \; . \; . \; . \quad (11)$$

We may therefore calculate δ directly from (11) if we know the relation between the initial value $K\,(=U)$ of the kinetic energy and the impulse P applied. To find this we consider that the initial velocity v at the ends of the rod will be zero and at the middle a maximum (vmA). If we assume that at the middle the impulse is completely converted into movement then the following is valid (m is the total mass of the rod):

$$\frac{m}{l}\,v_{\max} = \int p\,dt = \frac{P}{l},$$

and with a sinusoidal distribution of the initial velocity (corresponding to a sinusoidal form of the bent rod) the kinetic energy becomes

$$K = \int_0^l \frac{1}{2}\,\frac{m}{l}\left(v_{\max}\sin\frac{\pi x}{l}\right)^2 dx = \frac{m}{4}\,v^2_{\max},$$

thus $K = \dfrac{1}{4m}\,P^2$.

The impulse $P = \int pl\,dt$ may be calculated from equation (10), where for the sake of simplicity one may integrate from zero to infinity, since the "tail" of the force curve does not furnish any appreciable contribution. The result of the integration is:

$$P = 10.0 \text{ dyne sec.}$$

Thus $\quad U = K = 20.2$ ergs

and according to (11): $\quad \delta = 0.0105$ cms.

A more exact calculation gives a value which agrees very well with this ($\delta = 0.0117$ cms, see below).

The deviation of the wires is thus reduced to a

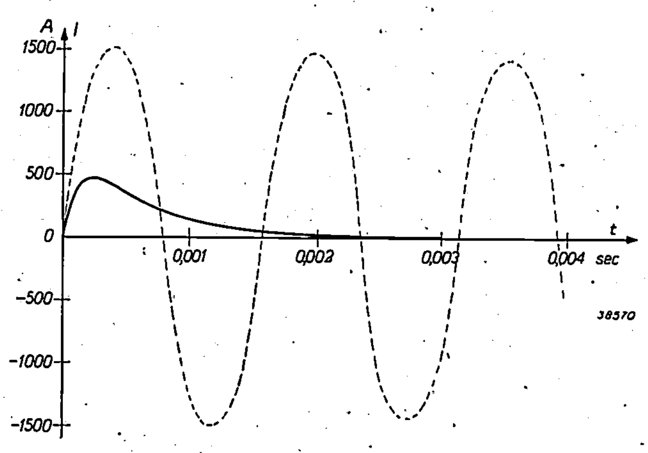


Fig. 9. Behaviour of the current i with a discharge similar to that in fig. 3, but with a resistance of 20 ohms in place of that of 0.1 ohm. The discharge is now aperiodic and practically dies out after 0.002 sec. The scale of the abscissa is 10 times that used in fig. 3 For the sake of comparison the behaviour of the current in fig. 3 is indicated by a broken line.

fraction of the original value (equation (8)) by the connection of the resistance of 20 ohms in the anode connections. The highest tensile stress in the wire becomes only about 100 kg/cm² in this case and no permanent deformations result.

Appendix: Exact calculation of the deviation

If x is the coordinate along the rod, u the transverse deviation which depends upon x and the time t, the following differential equation holds:

$$\frac{\partial^2 u}{\partial t^2} + c^2 \frac{\partial^4 u}{\partial x^4} = \frac{p(x,t)}{\varrho q}, \quad \ldots \ldots \quad (12)$$

where
$$c^2 = \frac{EI}{\varrho q}$$

and where the initial conditions are

$$u(x,0) = 0 \text{ and } \left(\frac{\partial u}{\partial t}\right)_{t=0} = 0.$$

In our case, where the force p depends only on t and not on x, the solution of the differential equation may be written in the form [5])

$$u(x,t) = \sum_{n=0}^{\infty} f_n(x)\varphi_n(t). \quad \ldots \ldots \quad (13)$$

[5]) The general solution which also holds for the case where p depends upon x is found in H. Schmidt, Zur Theorie der erzwungenen Transversalschwingungen homogener Stäbe konstanten Querschnitts, Z. Phys. **64**, 411, 1930.

The total vibration form is here built up of the so-called natural vibration forms

$$f_n(x) = \sin\frac{(2n+1)\pi x}{l}, \quad \ldots \ldots \quad (14)$$

each multiplied by a certain time function:

$$\varphi_n(t) = \frac{4l^2}{\pi^3 \varrho q c}\,\frac{1}{(2n+1)^3}\int_0^t p(s)\,\sin\frac{(2n+1)^2\pi^2\,c(t-s)}{l^2}\,ds\,. \quad (15)$$

In (15) the factor $1/(2n+1)^3$ occurs which decreases rapidly with increasing n, in addition to which the integration also gives a factor of the order of $1/(2n+1)^2$. We may therefore neglect the contribution of the higher natural vibration forms to the total vibration, i.e. the rod moves practically according to its fundamental vibration ($n = 0$), whereby according to equation (14) it takes on the form of a half sine. The solution (13) now becomes simply

$$u(x,t) = \frac{\sin\pi x}{l}\,\varphi_0(t),$$

where the time function $\varphi_0(t)$ for our example (with $R = 20$ ohms) can be calculated in an elementary way from (15) when the expression (10) for the force is substituted in it. The maximum deviation δ in which we are interested occurs in the middle of the rod in the fundamental vibration, where $\sin \pi x/l = 1$, so that δ becomes equal to the maximum value of $\varphi_0(t)$. By setting the differential quotient $d\varphi_0(t)/dt$ equal to zero it is found that the first maximum occurs at $t = 0.0025$ sec, and that δ then has a value of 0.0117 cm.

A METHOD OF MEASURING IN THE INVESTIGATION OF BICYCLE DYNAMOS

by H. A. E. KEITZ. 621.313.322 : 629.118.3

If we consider an A.C. dynamo with constant magnetic field, connected to an ohmic resistance, and plot the terminals voltage E as a function of the number of revolutions n, we obtain a curve of the type sketched in *fig. 1.* The voltage at first increases proportionally with n and reaches a constant value at higher values of n. The region of the characteristic on which the dynamo will work when in use depends not only upon its construction but also very much upon the resistance to which it is connected.

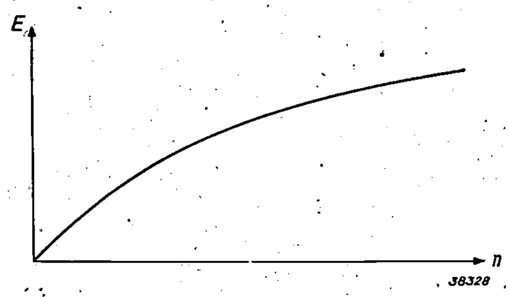


Fig. 1. General form of the variation of the terminals voltage E of an A.C. dynamo loaded with a resistance, as a function of the rate of revolution n.

In the case of bicycle dynamos the aim will be to have the operating point in ordinary use lie on the flat part of the characteristic as far as possible. The rate of revolution in the case of the bicycle dynamo varies very widely, since it is practically proportional to the speed of the bicycle, which may vary for instance from 3 to 20 m.p.h. If the characteristic is not sufficiently flat in the region corresponding to these speeds, it means that the lamp connected to the dynamo is already very much overloaded at speeds slightly higher than the normal, and will therefore quickly succumb, while at speeds slightly less than normal it gives hardly any light.

We shall not at this moment go into the structural measures by which the desired characteristic can be obtained [1]), but shall explain the important influence of the resistance in connection with the shape of the characteristic at normal rates of revolution. This could be demonstrated by plotting the relation between terminals voltage and speed of the bicycle for a number of different resistances. In practice, however, the dynamo is not loaded with a constant resistance, but with a lamp whose resistance varies quite sharply with the temperature of the filament and thus also with the terminals voltage. It is therefore better to study the relation between the terminals voltage and the speed of revolution for a given lamp.

In *fig. 2* this relation is given for the Philips bicycle dynamo, type No. 7 405 in connection with different lamps. It may be seen that with the lamp of 6 V-0.5 A, for which the dynamo is designed, the terminals voltage varies from 3.3 to 7.7 V for the two extreme speeds of the bicycle. With a lamp of 6 V-0.4 A the corresponding variation is from 3.8 to 10.7 V, with a lamp of 6 V-0.6 A it is 2.9 to 5.7 V. The lamp of 6 V-0.4 A, being too small, already reaches its nominal voltage at the low speed of 5 p.m.h., and at higher speeds therefore it will quickly succumb; the too large lamp of 6 V-0.6 A does not burn at the nominal voltage (and with the nominal light flux) even at the highest speeds of 20 p.m.h.

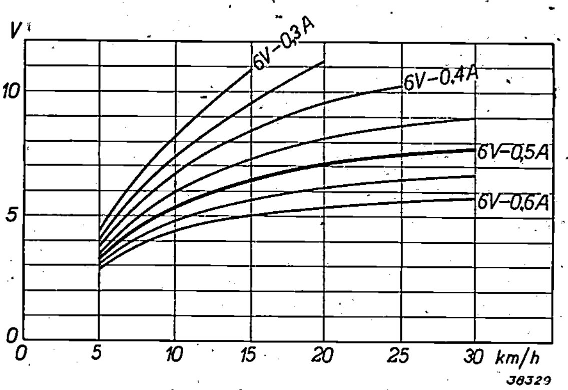


Fig. 2. Terminals voltage of the Philips bicycle dynamo 7 405 as a function of the speed of the bicycle for a load consisting of one of a number of different lamps. The best compromise between length of life and light flux of the lamp is obtained with the characteristic indicated by a thick line.

This shows clearly that dynamo and lamp must be mutually adapted [2]).

The testing of this adaptation amounts to recording the characteristic of the dynamo, preferably under normal conditions of use. Special attention must hereby be paid to two points. The load on the dynamo may not be changed by the connection of the measuring instruments, since the charac-

[1]) See on this subject: H. A. G. Hazeu and M. Kiek, An alternating current dynamo with a flat characteristic for bicycle illumination, Philips techn. Rev. 3, 87, 1938.

[2]) If two lamps in parallel are connected to the dynamo, the total current must have the prescribed value. If for example there is a lamp for 4-6 V-0.04 A in the rear light of the bicycle, a lamp of 6 V-0.45 A must be chosen for the headlight if the same characteristic is to be obtained as for a 6 V-0.5 A lamp alone.

teristic is so very sensitive to such a change, *i.e.* the measurement of the voltage must require practically no energy. Furthermore the A.C. voltage furnished by bicycle dynamos is in general not truly sinusoidal, so that account must be taken of whether or not the voltage measured actually corresponds to the effective value.

A very simple method of measuring, which avoids all difficulties in this respect and which we have now used for some time, is the following. The lamp which represents the load on the bicyle dynamo is connected successively to the dynamo and to an accumulator battery with a variable series resistance. If this resistance is so adjusted that the lamp gives the same light flux in both cases, the voltage on the lamp will then also be the same in both cases. The measurement of the terminals voltage of the dynamo is thus replaced by a simple D.C. voltage measurement which can be carried out with great accuracy. In order to be able to adjust the lamp to equal light flux, it is placed in a photometer sphere, *fig. 3* (instead of a sphere a box of any desired form may be taken) at the measuring window of which a photoelement is placed (a blocking-layer photocell). The photocurrent, which is read off on a milliammeter, is a measure of the light flux.

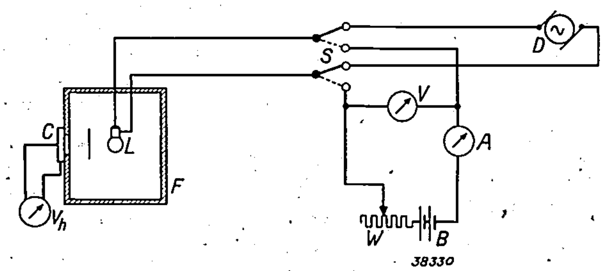


Fig. 3. Arrangement for the recording of dynamo characteristics. The dynamo D is driven by a motor with a variable number of revolutions which may be read off with a tachometer. The lamp L which is placed in a "photometer sphere" F is connected alternately to the dynamo and to an accumulator battery B by means of the switch S. The series resistance W is so adjusted that the lamp gives the same light flux in both cases, which is checked by means of the milliammeter V_h which indicates the photocurrent of the blocking-layer photocell C. The lamp voltage measured with the voltmeter V is then equal to the effective value of the terminals voltage of the dynamo. If desired, the lamp current can also be controlled with A.

The D.C. voltage measured will be equal to the effective value of the A.C. voltage supply, if it is permissible to assume that the light flux of the lamp is determined by this effective value, and no longer depends upon the form of the A.C. voltage. This is indeed the case if the frequency of the A.C. voltage is sufficiently high. Experience shows that even at the ordinary mains frequency of 50 c/s there is no measurable deviation of the light flux compared with that upon supply be means of a D.C. voltage equal to the effective value. Since the frequency of the A.C. voltage furnished by bicycle dynamos is in general considerably higher than 50 c/s — in the case of the Philips dynamo 7 405 it is already 70 c/s at a speed of 3 p.m.h. — the light flux may immediately be used as intermediate in the voltage measurement. Furthermore the requirement that no extra load on the dynamo may result from the measurement is here automatically satisfied.

The accuracy of the method is more than adequate. The voltage measurement for each point of the characteristic to be measured (*i.e.* for each rate of revolution) requires three readings, namely two on the milliammeter for the light flux and one on the voltmeter for the lamp voltage. The readings of the light flux are not, however, critical, since with the lamps used here the light flux varies about 3.5 times as much proportionally as the supply voltage, so that an inaccuracy in the adjustment to equal light flux has only a small effect.

It might be imagined that the comparison of the dynamo with the battery could be made superfluous by calibrating the milliammeter V_h directly in volts for a given lamp. Instead of three readings only one reading would then be necessary. This immediately meets with the objection that the ratio between light flux and voltage of the lamp changes with time: especially upon the recording of measured points in the region of high speeds (high voltages) a relatively rapid blackening of the bulb of the lamp takes place which would cause the calibration to change gradually. In the above-described comparative measurements this of course causes no difficulty if the readings are carried out in quick enough succession.

$\omega L = \frac{1}{\omega C}$

$\frac{\omega L}{2}$ $\frac{1}{\omega C r}$ $Q = \frac{L}{Cr}$

RESONANCE CIRCUITS FOR VERY HIGH FREQUENCIES

by C. G. A. von LINDERN and G. de VRIES. 538.565

Different concepts connected with the properties of electrical resonance circuits are discussed in this article. In particular the reasons are given why, in radio technology, from a coil and a condenser there has been a development toward cavity resonators, Lecher systems and cavity resonators for higher and higher frequencies.

Introduction

A familiar property of Maxwell's equations may be described in the following way. If all the dimensions of a system (and also the specific resistance ϱ) are reduced n times, the same voltages and currents will occur in this system as in the original one, if the frequency is multiplied by n (the wave length is n times smaller). It might be supposed that this would furnish sufficient guidance for the construction of short wave apparatus, and particularly of resonance circuits for short waves, were it not for the fact that for all kinds of reasons it is often impossible to decrease all the dimensions proportionally. Especially the reduction of the specific resistance in the desired proportion is often impossible since copper is used for the conductors even at low frequencies, and there is no available material which conducts appreciably better. We must therefore devise constructions which exhibit considerably better properties than the coils and condensers ordinarily used at low frequencies. Such constructions do actually exist; they are not, however, used at low frequencies because their dimensions would be too large for that purpose. It is because of the fact that at high frequencies all the components are proportionally smaller that these relatively large dimensions are then no objection. While of course the properties of these circuits also become worse upon reduction in size, due to the fact that the conductivity cannot be increased, nevertheless, because they are so much better than the ordinary ones at low frequencies, they are still just good enough at high frequencies.

We shall attempt to discover the line which has been followed in this development. Before doing so we shall first review briefly a few of the concepts connected with resonance circuits in general.

Resonance width and quality factor

The simple LC circuit with which we begin is represented generally as a connection in series of a resistance r and a self-induction L, in parallel to which is connected a capacity C (*fig. 1*). The total impedance Z of such an oscillating circuit is given by the formula:

$$\frac{1}{Z} = \frac{1}{r + j\omega L} + j\omega C. \qquad (1)$$

By resonance frequency we understand the frequency for which Z is a maximum. With an ohmic

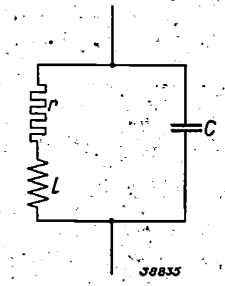


Fig. 1. Simple oscillating circuit consisting of resistance r, self-induction L and capacity C.

resistance r which is small compared with the reactances ωL and $\frac{1}{\omega C}$, the following is approximately valid for this frequency:

$$\omega_0^2 LC = 1. \qquad (2)$$

The total impedance Z of the LC circuit is real upon resonance, and according to (1) and (2) can be represented by

$$R = \frac{L}{Cr}. \qquad (3)$$

For an angular frequency which differs by $\Delta\omega/2$ from the resonance frequency, let the impedance have fallen to $R/\sqrt{2}$ (*cf. fig. 2*). For $\Delta\omega$ one then finds

$$\Delta\omega = \frac{r}{L}(\text{if } r \ll \omega L). \qquad (4)$$

This quantity $\Delta\omega$ is called the resonance width of the circuit, and the quotient of the angular frequency ω_0 at resonance and this resonance width $\Delta\omega$ is a measure of the sharpness of resonance of the circuit and is called the quality factor Q:

$$Q = \frac{\omega_0}{\Delta\omega} \text{ and with (4), } = \frac{\omega_0 L}{r} \qquad (5)$$

The resonance is sharper and the quality factor greater, the smaller the ohmic resistance r and the greater the self induction L. Under those circumstances a freely oscillating circuit also has thc least damping. This realtion between the sharpness of resonance for a forced vibration and the damping of a free vibration will be considered somewhat more closely.

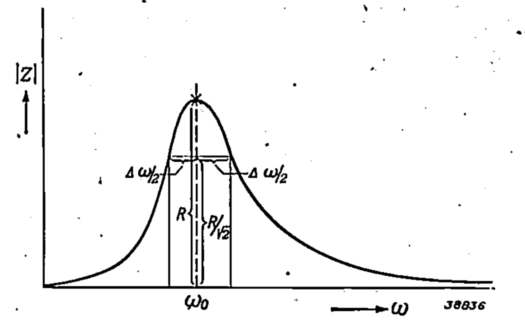


Fig. 2. Variation of the absolute value $|Z|$ of the impedance of a connection in parallel of a capacity C and a self-induction L as a function of the angular frequency ω. At the resonance frequency ω_3 the impedance is practically equal to L/Cr, where r represents the resistance.

Quality factor and damping

The frec oscillation in an electric circuit with a characteristic oscillation time T may be written as follows:

$$i = i_1 \sin \frac{2\pi t}{T} = i_0 e^{-\Theta\, t/T} \sin \frac{2\pi t}{T} \quad \cdots \quad (6)$$

In this expression, if Θ is small enough, i_1 represents the "momentary" amplitude of the slowly decreasing oscillation (*fig. 3*). The quantity Θ is called the logarihmic decrement and represents the natural logarithm of the ratio between two successive amplitudes. With an ohmic resistance r the heat losses during one period T are equal to $^1/_2\, i_1^2\, rT$. The total energy content of the oscillating circuit is $^1/_2\, L\, i_1^2$ and it decreases proportionally with

$$[e^{-\Theta\, t/T}]^2 \approx 1 - \frac{2\Theta t}{T}.$$

The decrease per period is thus $\Theta L\, i_1^2$ and this must be equal to the heat $^1/_2\, i_1^2\, r\, T$ developed per period. It therefore follows that the logarithmic decrement

$$\Theta = \frac{rT}{2L} \quad \cdots\cdots\cdots \quad (7)$$

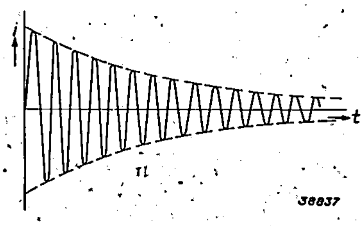


Fig. 3. Damped vibration.

Since $\omega_0 T$ is equal to 2π, the following relation is found between logarithmic decrement and quality factor

$$\Theta = \frac{\pi}{Q} \quad \cdots\cdots\cdots \quad (8)$$

If we writc formulac (7) and (8) in the form

$$Q = \frac{2\pi}{T} \frac{L i_1^2}{r i_1^2}, \quad \cdots\cdots \quad (9)$$

which is also in accordance with (5), we have obtained Q in a form which is also significant when we are no longer concerned with a simple LC circuit. Equation (9) is then read as follows:

$$Q = 2\pi \cdot \frac{\text{field energy}}{\text{energy dissipated per period}}; \quad (9a)$$

in many cases for more complicated systems (thc cavity resonators, resonating cavities and Lecher systems to be described) it can be proved that equations (8) and (9a) are also valid.

An exception is formed by band-pass filters for example. Under certain circumstances there may be several resonance frequencies, two or more of these frequencies may not, however, lie too close to each other, as is the case in the band-pass filter which is used in the intermediate-frequency amplifiers of a radio receiving set.

Skin effect

For the resistance r occurring in the formulae we must not use the ordinary D.C. resistance but the A.C. resistance which is so much higher due to the well known skin effect. We shall not here go into the theory of the skin effect, but shall only recall a few main features of it [1]). At very high frequencies, no magnetic field — or strictly speaking only a very weak one — will be found in the interior of a good conductor. This is understandable when it is kept in mind that due to the high frequency the result of a magnetic alternating field would be that high voltages would be induced in the conductor, which would lead to strong eddy currents of such a sort that they would cause in turn a magnetic field which is oppositely directed to the original field. In a straight wire of circular cross section at very high frequencies the current flows in a thin layer on the outside of the wire. The fact that this situation satisfies the requirement of producing

[1]) The phenomenon of skin effect should not be conceived as if the currents repelled each other because:
1. conductors through which currents flow in the same direction do not repel but attract each other,
2. the forces act between the conductors and not between the currents as such,
3. there is no skin effect for direct current.

no magnetic field in the interior is due to the fact that a cylindrical ring of uniformly distributed current lines of force causes a magnetic field toward the outside, but none toward the inside (*fig. 4*).

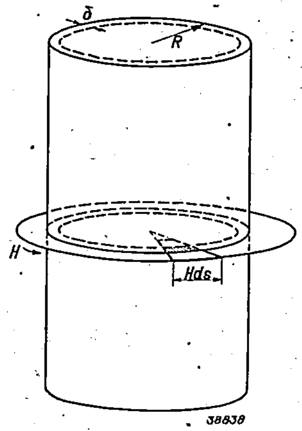


Fig. 4. Diagrammatic representation of a straight conductor of circular cross section in which skin effect occurs, so that the current flows chiefly in the layer δ.

The line integral of the magnetic force $2\pi RH$, which must be proportional to the "enclosed" current, will be zero when we choose our radius smaller than that of the current lines cylinder, since no current is enclosed on the inside.

The result of the skin effect is that only a part of the cross section of the conductor is used for the current, which amounts to an apparent increase in the resistance. The current density in the current carrying layer decreases from the outside inwards as $e^{-x/\delta}$, where x represents the depth below the surface of the conductor (*fig. 5*). In the case of a wire of circular cross section, this is therefore in the direction of the radius. δ is called the depth of penetration. The current becomes very small in the interior of the conductor, but not exactly zero, in accordance with the e power. It may still be asked how thick the wall of a hollow conductor must be to develop the same amount of heat as the solid conductor with a given total current in both cases. This might be called the equivalent thickness. It is then found that the equivalent thickness is equal to the depth of penetration [2]).

For δ is found:

$$\delta = c_1 \sqrt{\varrho \lambda},$$

where ϱ is the specific resistance and λ the wave length. The fact that the layer is thicker with higher specific resistance is also quite easily understandable: the mechanism which keeps the interior of the conductor free of field acts less perfectly in a poor conductor than in a very good conductor, due to the ohmic losses. For copper the depth of penetration becomes:

$$\delta_{\mathrm{cm}} = 4 \cdot 10^{-5} \sqrt{\lambda_{\mathrm{cm}}} \quad . \quad . \quad . \quad . \quad (10)$$

For the resistance of a conductor of circular cross section one finds:

$$r_{\mathrm{w}} = \varrho \frac{l}{2\pi R \delta} \quad . \quad . \quad . \quad . \quad (11)$$

instead of, as in the D.C. case:

$$r_0 = \varrho \frac{l}{\pi R^2} \quad . \quad . \quad . \quad . \quad (12)$$

Furthermore for a flat strip (see *fig. 6*):

$$r_{\mathrm{w}} = \varrho \frac{l}{2(b+a)\delta} \quad . \quad . \quad . \quad . \quad (13)$$

If the frequency is not too high, so that the current is no longer concentrated in a very thin layer, the formulae are much more complicated [3]) than the simple exponential law mentioned above. Such cases are, however, of no importance for our subject, resonance circuits for very high frequencies.

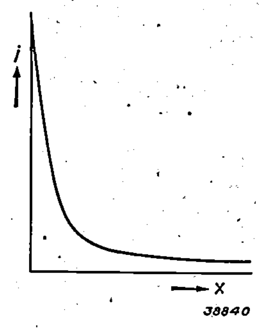


Fig. 5. Current distribution over the cross section of an electric conductor due to the skin effect. The current density i varies approximately exponentially with the distance x to the surface of the conductor.

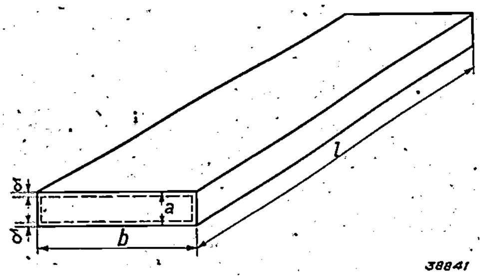


Fig. 6. Conducting wide strip of a length l, a width b and a thickness a. The depth of penetration for a high-frequency alternating current amounts to δ because of the skin effect.

If there is more than one straight conductor,

[2]) In the derivation the phase shifts occurring in the layer must be taken into account.

[3]) See: H. G. Möller, Grundlagen und mathematische Hilfsmittel der Hochfrequenztechnik (Springer, Berlin 1940), who on page 34 also begins with a coil of wide strip and considers particularly the lower frrquencies.

or if for other reasons a magnetic A.C. field is already present, the current density is not the same at every spot on the surface of the conductor, but a current distribution occurs such that the field in the interior of the conductor is again very small. In order to ascertain on which side of the surface of the conductor the most current will be encountered, use may be made of the fact that the current on the surface is greatest where the magnetic field is greatest.

Since the magnetic field within the current layer is negligibly small, the line integral in fig. 4 is equal to Hds (the two sides are perpendicular to H and thus do not contribute to the line integral); the current enclosed becomes $i_1 ds$ when there is a current i_1 per centimetre circumference of the layer, so that H is proportional to i_1. This rule is only useful in cases where the skin effect occurring does not alter the shape of the magnetic field outside the conductor too much, but such cases frequently occur.

Thus in the case of a cylindrical coil the greatest current density will be found at the inside of the coil where the magnetic field is strongest (*fig. 7*). Practically no current flows on the outside in the case of long coils.

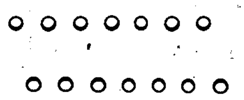


Fig. 7. Diagram of the current distribution in the wires of a cylindrical coil as a result of the skin effect.

On the basis of a consideration of the coil made of wide strip — which somewhat unusual type of coil is chosen because of the simplicity of the calculations — and the torus, we shall now attempt to explain why high-frequency technique, beginning with the ordinary coil has developed in the direction of the c a v i t y r e s o n a t o r, L e c h e r system and c e l l - r e s o n a t o r which are always more important on short waves.

The long coil made of wide strip

In a long straight conductor (fig. 6) with a length l, a width b and a thickness a which is small compared with b, a current of sufficiently high frequency flows mainly through a surface layer of the thickness δ due to the skin effect just discussed, so that the resistance is

$$r_\mathrm{w} = \varrho \frac{l}{2b\delta} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

This long wide strip is now bent to form a coil of one turn (*fig. 8*), forming thus a cylinder with a thickness a, a circumference $l = 2\pi R$, when R represents the radius, and a length b. The result is that current now flows only on the inside of the cylinder (with the depth of penetration δ), as indicated in fig. 8, so that the resistance of such a

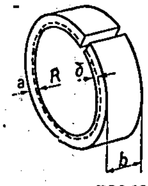


Fig. 8. The length of wide strip of fig. 6 is bent to one winding with a circumference $l = 2\pi R$, a width b and a thickness a. As a result of the skin effect high-frequency alternating currents now flow only through a layer δ lying on the inside of the winding.

coil of wide strip for sufficiently high frequencies is twice as great as the A.C. resistance for the straight wide strip of which the coil is made. The A.C. resistance for the strip bent to a coil of one winding is therefore

$$r_\mathrm{w} = \varrho \frac{l}{b\delta} = \varrho \frac{2\pi R}{b\delta} \quad . \quad . \quad . \quad . \quad (15)$$

If the strip is wound to a coil of n turns (*fig. 9*) the same considerations are valid, and the A.C. resistance becomes:

$$r_\mathrm{w} = \varrho \frac{2\pi R}{b\delta} n. \quad . \quad . \quad . \quad . \quad . \quad (16)$$

The self-induction, when the coil is long enough with respect to its width, is:

$$L = 4\pi^2 R^2 \frac{n}{b} \cdot 10^{-9} \text{ henrys.} \quad . \quad . \quad (17)$$

From formulae (16) and (17) for resistance and self induction of a long coil of wide strip, its quality factor follows directly:

$$Q = \frac{\omega_0 L}{r_\mathrm{w}} = 2\pi R \frac{\omega_0 \delta}{\varrho} \cdot 10^{-9} \quad . \quad . \quad . \quad (18)$$

Since δ is proportional to $\sqrt{\varrho/\omega_0}$ the quality of the coil is inversely proportional to $\sqrt{\varrho}$ and directly proportional to $\sqrt{\omega_0}$. The latter fact seems to offer hope for the high frequencies, but this is not the case. It means that for a given coil the quality be-

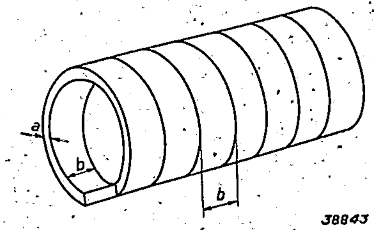


Fig. 9. Straight coil of wide strip, consisting of n windings with a radius R, a width b and a thickness a.

comes better at high frequencies — thus smaller tuning capacity — but the difficulty is that for all kinds of reasons it is often impossible to go lower than a certain tuning capacity so that smaller self inductions are arrived at. This means according to (7) a decrease of the radius R with increasing frequency, whereby according to (18) the quality factor becomes poorer. This effect dominates over the increase in Q with $\sqrt{\omega_0}$ just discussed, so that Q actually decreases with increasing frequency.

In order to allow the radius R to remain as large as possible at high frequency (small self-induction), it is advisable to use a coil with relatively few wide windings. A new source of losses then occurs, however: the coil begins to resemble a loop aerial more and more and will give off energy by radiation, which decreases the value of Q in exactly the same way as the ohmic losses. The torus coil — which we here include in the discussion exclusively as a transition to the cavity resonator meets this objection, and is sometimes actually employed for this reason.

Strictly speaking, it is not usually a question of radiation of tuned circuits, but of induction in more or less poor conductors in the neighbourhood. Another way besides the one here described of decreasing these losses consists in placing the coil in a box of conducting material.

The torus coil

The torus coil is formed from the ordinary cylindrical coil by bending it into a circular ring (*fig. 10*), so that the magnetic lines of force are continuous inside the coil. As a result of this the formulae (16), (17) and (18), which were derived neglecting the radiation losses for the long straight coil, are exactly valid for this case. In order to obtain a good Q even at high frequencies in spite of the small self-induction, we may therefore make the windings large but small in number, without introducing appreciably large extra losses. Proceeding in this way, fewer and fewer windings are used at higher frequencies, until finally only one remains

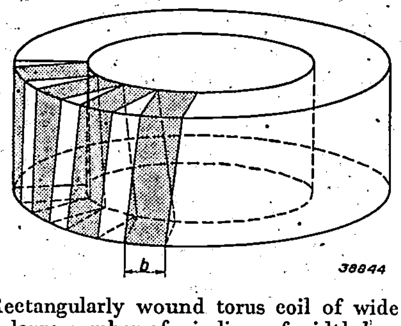


Fig. 10. Rectangularly wound torus coil of wide strip, consisting of a large number of windings of width b, within which the magnetic field is enclosed.

which is just as wide as the circumference of the torus (*fig. 11*). We then have a torus surface which is cut open along a circle in its upper surface.

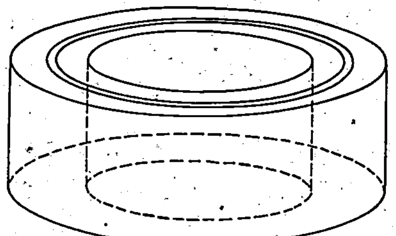


Fig. 11. Rectangularly wound torus coil of one winding with a width equal to the entire circumference of the torus. The coil is cut open on its upper surface along a circular line.

The special form of self-induction L must now be completed with a capacity to form a high-frequency oscillating circuit. If we do this in the manner represented diagrammatically in *fig. 12*, we have derived the so-called cavity resonator from the torus coil.

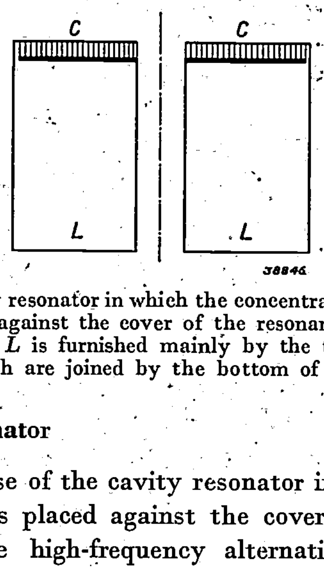


Fig. 12. Cavity resonator in which the concentrated capacity C is introduced against the cover of the resonantor, while the self induction L is furnished mainly by the two concentric cylinders which are joined by the bottom of the resonator.

Cavity resonator

In the case of the cavity resonator in fig. 12 the capacity C is placed against the cover of the resonator. The high-frequency alternating currents to and from the condenser plates flow through the axis, the bottom and the inside of the outer wall. The electrical and magnetic fields are both therefore inside the resonator.

It may in general be said of the quality factor of such a circuit that it is proportional to the quotient of volume and surface of the resonator. Although the cavity resonator is not large compared with the wave length, this is entirely analogous to Sabine's law about the reverberation time in a room. This reverberation time in a room which is large compared with the wave length of the sound is proportional to the quotient of the volume and the sound absorbing surface. The reverberation time is the damping time and therefore inversely proportional to the logarithmic decrement, *i.e.* proportional to the quality factor Q according to formula (8). Formula (18) for the quality factor of a torus coil is in complete agreement with this general statement, because according to that formula Q is proportional to the radius R of the torus. i.e. to the quotient of volume and surface.

Meanwhile it must be kept in mind that this statement only forms a guide and must be used with care. The proportionality factor also contains certain other quantities which take into account the shape of the space, especially when it is small with respect to the wave length. It therefore serves no useful purpose to increase the volume by making the core thinner, on the contrary, the higher ohmic resistances which will thereby be caused will quickly make the value of Q considerably worse. It can, however, be said that long thin resonator models are poorer than more or less "square" ones, since the latter have a much more favourable relation between volume and surface [4]).

In order to reach a high impedance it is of still greater importance than for attaining very good quality to make the self-induction large, since $Q = \omega L/r$ and $R = L/Cr = \omega^2 L^2/r$. This is why in cases where one is not bound for some reason or other to a minimum capacity, the condenser of the cavity resonator must be made as small as possible. The introduction of a specially concentrated capacity as in fig. 12 will then finally be omitted, and only a more or less shortened core is retained. If its length is fairly large compared with the diameter of the resonator the system becomes that of two concentric conductors, which is usually called a concentric Lecher system, *fig. 13*. The dimensions of our "circuit" are now no longer small compared with the wave length, on the contrary, the resonance wave length is found to be only about four times as great as the length of the "core" in fig. 13, as follows from the theory of Lecher systems, to which we shall return in a coming number of this periodical.

The attempt to increase still further the ratio between volume and surface leads to increasing the diameter of the concentric Lecher system. The relative thinness of the core, however, stands in the way of any considerable improvement in Q due to its high ohmic losses.

It is found that when we have increased the diameter of the resonator until it has the same order of magnitude as the wave length, we then still find resonances when the core has entirely disappeared. Such a system is called a resonance cavity (*cf. fig. 14*).

We have then, however, adopted a whole new group of concepts in our considerations. From the moment when we decrease the capacity of the cavity resonator and therby increase the self-induction so much that all the dimensions are no longer small with respect to the wave length, the phenomena are no longer quasi-stationary.

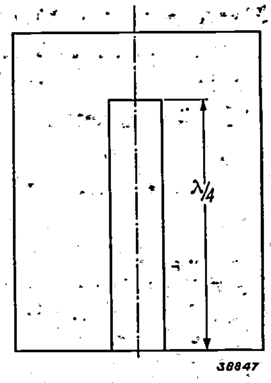


Fig. 13. Concentric Lecher system consisting of a closed metal cylinder with an empty core projecting inwards. The system resonates for a wave length λ which is about four times the length of the core.

Quasi-stationary and non-quasi-stationary

Stationary, quasi-stationary and non-quasi-stationary phenomena may be distinguished. Stationary phenomena are spoken of in the case of systems which are at rest (electrically and magnetically). They are still considered to be at rest when a direct current flows. In this case concepts occur such as resistance, capacity and self-induction, which are defined perfectly unambiguously, and the current in a satisfactorily insulated conductor is the same at all points. In quasi-stationary phenomena the state of rest is disturbed; the currents and voltages are not the same at each moment. The current in a conductor is indeed still the same through its whole length, and as long as all changes take place slowly enough, concepts such as capacity self-induction and resistance retain their significance. In the equations which describe the phenomena which take place in this region, total differentiations with respect to time occur.

What must be understood by "slowly enough" becomes clear upon studying the phenomena which are indicated as non-quasi-stationary. The so-called displacement current which leaves the conductors laterally now begins to play a part, so that the current through the conductor is no longer the same at all points. The equations begin to contain

[4]) If one studies in detail what is the best model one does not find that the diameter should be exactly equal to the length, but neither does one arrive at very long or flat models.

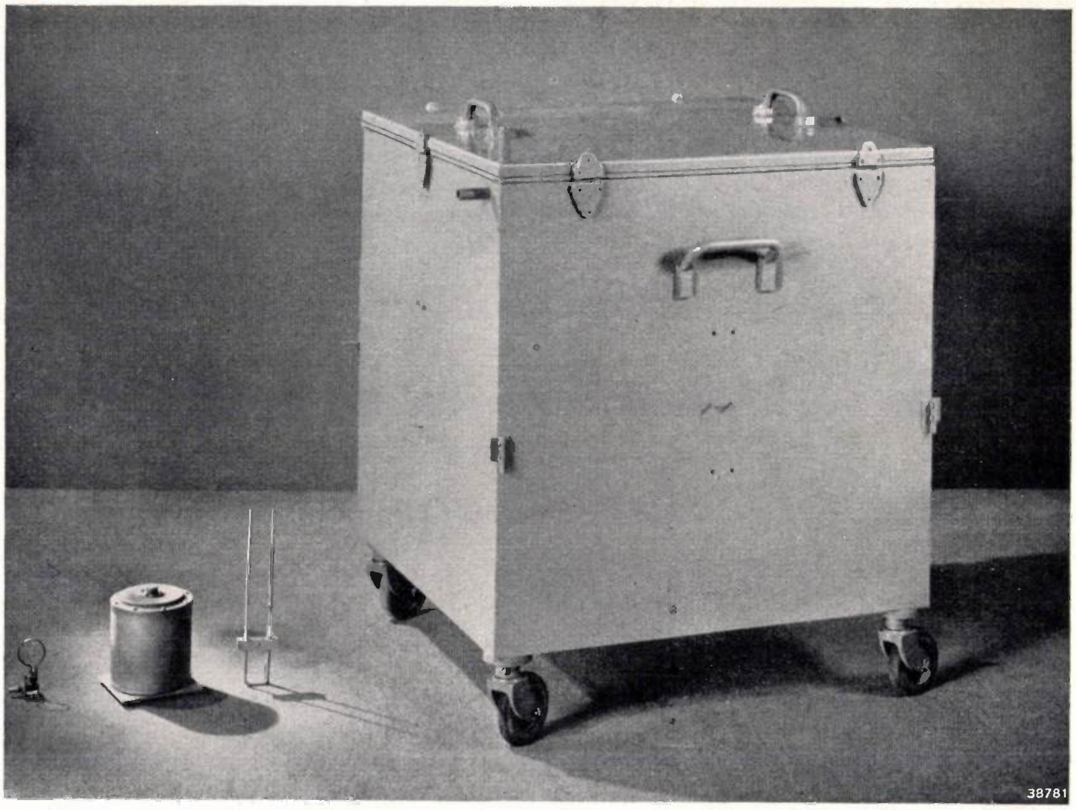


Fig. 14. An *L.C.* circuit, cavity resonator, Lecher system and resonance cavity, all of which are tuned to a wave length of about 1 m, are shown side by side for comparison of their dimensions in this photograph. In the middle of the front wall of the resonance cavity may be seen the connecting wires for the loop situated inside the cavity for the purpose of bringing it into resonance.

partial differentiations with respect to time, their solutions take on a wave character, and we are now also able to define more sharply the "slowly enough" of the foregoing paragraph. When the dimensions are no longer small with respect to the wave length the phenomena are no longer quasi-stationary. In cases in which one dimension is of the order of the wave length, while the others are much smaller, one might speak of semi-quasi-stationary. The Lecher system is an example of this.

In "semi-quasi-stationary" systems a certain significance may still be ascribed to the concepts capacity and self-induction, if they are considered for not too long sections of the system at once. In the entirely non-quasi-stationary systems, to which belong resonance cavities, there can be no question of this.

For the sake of completeness it must be noted that the criterion "large compared with the wave length in a vacuum" is not decisive in all cases. The skin effect for example, where the density and phase of the current depend closely upon position, is a non-quasi-stationary phenomenon, in spite of the fact that it takes place in a layer which is very much thinner than the wave length.

Lecher systems

In our survey of the different types of resonance circuits which are used in short wave technique we have encountered the concentric Lecher system but not the ordinary Lecher system. This is because we were chiefly concerned with the problem of keeping the value of Q satisfactory while passing to shorter and shorter wave lengths. In the meantime there are many cases where it is more a question of the impedance Z. This quantity, $Z_{\max} = L/Cr = \omega^2L^2/r$ may still be very satisfactory while $Q = \omega L/r$ is only moderately good, at least when L is large enough, which means that C should be able to be made small enough. In that case it is advantageous to use the ordinary Lecher system, which consists of two parallel conductors with a short-circuiting shunt which may be moved back and forth along the conductors. Although the Lecher system is not a quasi-stationary system, and one may not therefore speak of self-induction and capacity without some reservation, it is neverthe less clear that the system contains relatively little "capacity", and due to the fact that currents flow in opposite directions through the adjacent conductors, the radiation remains quite small without further shielding, so that the loss resistance does not become too high. The fact that the Lecher system has found such extensive application is due not only to the high value of Z_{max}, but also to the fact that resonance occurs when the length of the system is a whole number of quarter wave lengths, so that it can be used as a wave meter in a very simple way. We shall devote a subsequent article to the properties of the Lecher system, and shall therefore not go more deeply into the question here.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on applications to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabriken, Eindhoven (Holland), Kastanjelaan.

1535: W. Nijenhuis and F. L. Stumpers: On some properties of electrical networks (Physica **8**, 289, Feb. 1941).

The relation between the real or imaginary part of an impedance function and the function itself is derived, with special attention to transmitted impedances. As a result of these considerations two special types of electrical connections can be distinguished, namely those in which the modulus and those in which the phase of the impedance does not depend upon the frequency. The practical construction of such networks is discussed in the article. In conclusion several general properties are discussed of the variation of the phase with frequency.

1536: J. van Niekerk and M. S. C. Bliek: Het genezende effect van groote doses bestraald provitamine D van dierlijken oorsprong éénmaal, oraal of intramusculair, toegediend bij rachitis. (The curative effect of large single doses of irradiated provitamin D of animal origin, administered orally or intramuscularly, in the case of rickets). (Ned. T. Geneesk., **85**, 860-867, Mar. 1941).

As a continuation of the experiments described in **1532**, it is now shown that by administering a single large dose of vitamin D of animal origin to chicks orally or injecting it into a muscle, rickets can be cured. With the amounts used in these experiments no noticeable difference in action could be observed for the different methods of administration of the vitamin. The duration of the cure does not seem to depend upon the quantity used in nor upon the manner of administration. For insuring further normal calcification of the new bone tissue these conditions are, however, of importance. If the vitamin D is injected into a muscle a much smaller dose is sufficient when it is a question of preventing rachitic lesions after the cure is completed. If it is a question of curing rickets with a single dose of vitamin D, therefore, it is advisable to inject it into a muscle.

Contents of Philips Transmitting News **8**, No. 1, March 1941.

H. B. R. Boosman and R. P. Wirix, Transmitter and receiver tuning components.

5 kW tropic proof shortwave broadcast transmitter.

Tj. Douma, Resonance of circuits and lines.

VOL. 6. No. 8, p. 225-252 AUGUST 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

SMALL APPARATUS FOR MEDICAL X-RAY EXAMINATION

by H. A. G. HAZEU. 621.386.1 : 616.073

A description is given of the connections and construction of two small portable X-ray apparatus for medial diagnosis, the "Centralix" and the "Practix". The low-power appliances are used in cases where the patient to be examined cannot be transported to the X-ray laboratory of the hospital, and also for cases where it is unnecessary to use high intensities of X-rays (photography of the teeth, etc.). In such small apparatus the X-ray tube and the high-voltage generator are constructed as a single unit. In the "Centralix" apparatus this has been done by giving the iron of the high-voltage transformer the shape of a hollow cylindrical ring, in the cavity of which the windings are placed and in the axial core the X-ray tube. This gives a very compact and light unit. In the case of the "Practix", for the sake of a better heat dissipation, a less compact structure has been chosen; a heat balance shows that this apparatus may very well be used for continuous routine work (fluoroscopy). The two appliances may be used either without special supports or in combination with different specially constructed standards, some of which are portable.

The visitor to the X-ray laboratory of a modern hospital will generally be impressed by the large and seemingly complicated apparatus which he finds there: large X-ray tubes, supported by heavy standards, large high-voltage generators, heavy cables, complicated arrangements for placing the patient and the film holder in position, etc. [1]). The fact that such large apparatus has been found necessary in the development of X-ray technology is due to the desire to be able to obtain the best possible X-ray pictures in all cases, even for parts of the body which are difficult to photograph, such as lungs, stomach, pelvis, etc. For these purposes high X-ray intensities and thus high-power apparatus are required.

Because of its size and weight this apparatus is more or less bound to a permanent position and it is necessary to bring the patient to the apparatus in question and to place him in a certain position in relation to the apparatus. Often enough, however, cases occur in which this is impossible. It is only necessary to consider patients with serious fractures, with fractures to be treated by extension, patients who are confined to bed in their homes, etc. In order, nevertheless, to be able to make an X-ray examination in such cases, the doctor will be willing to accept a lower standard of quality in the X-ray picture or of universal applicability of his X-ray apparatus, if he can have an apparatus which is portable and which offers the necessary freedom in setting it up.

On the other hand there are also cases where ease in moving and adjusting the apparatus — although always an advantage — is not strictly necessary, but where only such low X-ray intensities are needed that the use of a high-power apparatus would be extravagant. This is true for instance of the X-ray examination of the jaw and teeth by the dentist.

For these cases in which either willingly or unwillingly the doctor must accept a lower X-ray intensity, Philips have been one of the first manufacturers to develop portable X-ray apparatus of low power (the first was the "Metalix" Junior apparatus which appeared on the market in 1927). At present mainly two appliances of this type are being made, and they will be described in the following.

General construction of X-ray apparatus

It is clear that appliances of high power will be large and those of low power small. But for a better understanding it is useful to point out the reasons

[1]) See for example H. A. G. Hazeu and J. M. Ledeboer. A universal apparatus for X-ray diagnosis, Philips techn. Rev. 6, 12, 1941.

for this on the basis of a survey of an X-ray installation.

The X-ray tube which works with voltages up to 100 kV in medical diagnosis, must in the first place be large enough to ensure the required high-voltage insulation, in particular that between the anode and the cathode leads. The dimensions become greater when the tube is surrounded by an earthed jacket for the protection of doctor and patient, since the necessary insulation distances must be maintained between jacket and leads. Furthermore, when the tube is intended for continuous use (fluoroscopy) it must also have a sufficiently large surface to dissipate the heat developed on the anode without excessive rise in temperature.

The size and weight of the high-voltage generator are determined, besides by the above mentioned requirements of insulation and heat dissipation, to a large extent by the desire to keep the heat development low, for which purpose a heavy core and heavy windings of the high-voltage transformer are necessary. At the same time this heavy construction of the transformer is also desired in order to limit the voltage losses which occur [2]) when large tube currents are taken off (up to 700 mA in large installations). In general the A.C. voltage, after having been transformed upwards, is rectified, since by this means the yield of X-radiation per mm^2 surface of the focus can be considerably improved.

In order to be able to adjust the X-ray tube sufficiently easily it is connected to the poles of the fixed generator by means of two long flexible high-voltage cables. As to the further accessories such as standards, regulation arrangements for voltage, current, time of exposure, etc., they become proportionally larger and more complicated as higher and higher requirements are made of the performance and universal applicability of the installation.

If our requirements are lower, then in the first place a lower tube voltage and a smaller tube current can be used. If the voltage is limited to 60 of 65 kV instead of 100 kV — it is impossible to go much lower since the efficiency of the excitation of the X-rays and the penetration of the rays becomes too small — tube and generator become considerably smaller, since the insulation difficulties increase much more rapidly than the voltage. The tube current can for many purposes be limited to 10 or 5 mA [3]); the high-voltage generator then becomes much smaller and lighter, especially when we do not make too great demands as regards the specific focal loading, and therefore omit the high-voltage rectifier with the necessary valves and other appurtenances.

When the tube and the source of high voltage have been reduced in this way to more modest dimensions it is only necessary to consider the cable, because the other parts of the installation have meanwhile disappeared automatically or have shrunk in size appreciably. The support for the patient is automatically eliminated in the applications for which the apparatus is intended, the standard for the tube becomes smaller and lighter in proportion to the tube itself, and the arrangements for regulation also become much simpler when we confine ourselves to less difficult objects and thus make fewer or no requirements as to adjustability of the voltage, etc. The dimensions of the high-voltage cables, however, and of their leads through the earthed covering of the tube are mainly prescribed by the necessary insulation, and have thus participated only to a limited extent in the shrinkage of the generator. While the cable was originally introduced in order to be able to adjust the tube easily free of the generator, it is now found that, from a given power value on, the cable itself becomes heavier and offers more hindrance to manipulation than the generator! It is obvious that having reached this limit of power it is better to omit the cable and combine the high-voltage transformer and the tube to a single unit. The necessary electrical energy can be supplied to the "unit" thus obtained from the light mains by means of a thin low-voltage cord.

This fundamental principle is realized in two different ways in the small Philips apparatus. We shall first consider the simpler apparatus, the "Centralix".

The "Centralix" apparatus

The connections

In the construction of this apparatus especial attention was paid to its application in the dentist's

[2]) This is of special importance for the fine regulation of the tube voltage; see in this connection the article referred to in footnote [1]).

[3]) In fact the quality of the X-ray picture need not immediately depreciate very much upon decrease in the tube current. A smaller tube current makes it possible, with the same maximum specific focal loading, to use a smaller focus (with a smaller focus the specific focal loading even becomes slightly greater due to the lateral heat dissipation in the anode block). Because of this for the same geometrical lack of definition, the distance between focus and object can be reduced, thus obtaining a greater X-ray intensity on the film. When the object is thick, however, an increasing distortion of the X-ray picture results. With thin objects, such for example as the jaw and teeth. hands, etc. this distortion is of no importance.

practice. In this application the same tube voltage of about 58 kV can always be used. Since in jaw exposures the distance between focus and film can be made very small (15 to 20 cm, with a focus of 1.2 × 1.2 mm, see footnote [3])) a tube current of 5 mA is sufficient to be able to work with reasonable exposure times (several seconds maximum) in all cases occurring.

While in apparatus with adjustable voltage and current a separate heating current transformer is necessary in order to be able to vary the tube current (*i.e.* the cathode emission) independently of the tube voltage, in this case, where only one current-voltage combination occurs and the heating voltage thus has a fixed relation to the tube voltage, the filament can be supplied from a pair of extra windings on the high-voltage transformer. The connections thus become extremely simple, see *fig. 1.* The secondary winding of the high-voltage transformer is earthed at the middle, so that each pole need be insulated for only half of the voltage (about 30 kV). At the cathode pole the extra windings S_3 for the heating current are added, as well as a small resistance R_1 with which the heating current is set at the correct value upon assembly of the whole. The primary winding of the transformer is connected to the 220 V mains *via* a switch which is operated by means of a timing device. With this timing switch, which may be compared with the shutter of a camera, the exposure time is regulated.

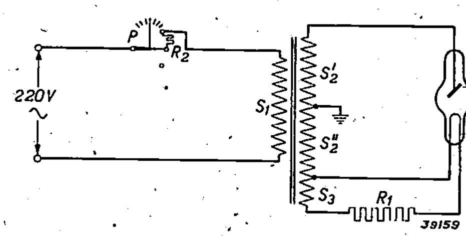


Fig. 1. Connections of the "Centralix" apparatus. B X-ray tube with anode A and hot cathode G; S_1 primary, $S_2' + S_2''$ secondary windings of the high-voltage transformer, S_3 extra windings for the filament current; R_1 regulation resistance for the heating current (regulated in the factory); P mains switch operated by a relay. Upon switching on, the resistance R_2 is first in series for several tenths of a second, and the voltage and current of the tube are slightly below normal, so that practically no X-radiation is excited, while the filament already nearly reaches the full required temperature.

The transformer

A nice solution has been found for the problem of combining the X-ray tube and the transformer. It is illustrated in *fig. 2.* Fig. 2*a* shows the simplest and most commonly used form of a transformer (so-called core transformer): the primary and secondary windings are wound around an iron core which is continued in the form of a yoke around the coils to give a closed magnetic circuit. The yoke can also be divided into two parts and the second half can pass around the other side of the

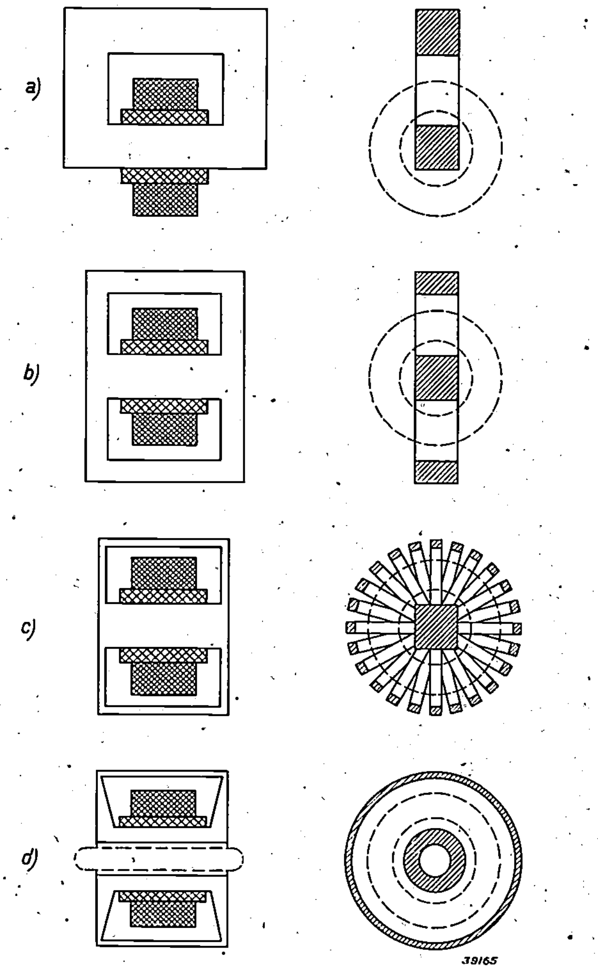


Fig. 2. By subdivision of the yoke of the ordinary core transformer (*a*) one arrives, *via* the shell transformer (*b*) at the cylindrical construction (*c*), from which by leaving a cylindrical opening at the axis the form of the "Centralix" apparatus is derived (*d*).

coils, see fig. 2*b* (shell transformer). In this case it is found that less iron is needed for a given induction. Still less iron is used when we go a step farther and divide the yoke into a large number of parts which surround the coils radially, see fig. 2*c*. In the "Centralix" apparatus this last method is actually followed, but with a space left free along the axis of the cylindrical structure formed. The X-ray tube is housed in this space, fig. 2*d*. In this way an extremely small and compact whole is obtained.

In *fig. 3* the construction of the transformer is given in somewhat more detail. The two halves of the secondary winding lie side by side as two separate coils around the cylindrical iron core, and the earthed extremities (*i.e.* the middle of the secondary winding) lie closest to the iron. In the successive layers of the winding the voltage becomes higher and higher and the outermost layers finally carry the full voltage of about 30 kV with respect to earth, so that here the two coils for about 30 kV with respect to the iron and for about 60 kV with respect to each other must be insulated. For the insulation between the coils and the iron a relatively small distance was already enough, because of the fact that the side of the iron facing the coils forms a completely smooth cylindrical surface with no corners or edges which might cause local increases of the field strength. In order, however, to be able to reduce the distance still further and thus make the iron body still smaller for the same volume of coil, each coil was surrounded by a "Philite" cylinder (C in fig. 3), while the intermediate spaces were poured full of a solid insulating mass (compound). In order to obtain the necessary mutual insulation distance between the tops of the two coils, the coils are wound with a cross section which becomes narrower toward the outside (trapezium-shaped). As may be seen in fig. 3, a V-shaped opening thus automatically occurs in the longitudinal cross section, which is necessary in order to allow the passage of the effective beam of X-rays from the tube lying inside. The X-rays emitted in other directions are cut off by the transformer, so that no special lead jacket is necessary for protection against these rays. This is an extra advantage of the method of construction here chosen with the tube and transformer combined, as is also the fact that the focus of the tube now lies close to the centre of gravity of the whole unit: because of this the focus remains in the same spot when the apparatus is suspended on a standard and rotated and this makes the adjustment easier.

The iron core is built up of laminae on planes through the axis, so that the diagram in fig. 2c of an iron circuit divided in many independent yokes is actually realized. The laminae do not fill the whole circumference but leave a small sector open (fig. 3) through which the leads D_1 and D_2 for the connections of the high-voltage winding are laid to the X-ray tube. At the same time this leaves the necessary opening in the transverse cross section for the passage of the effective X-ray beam.

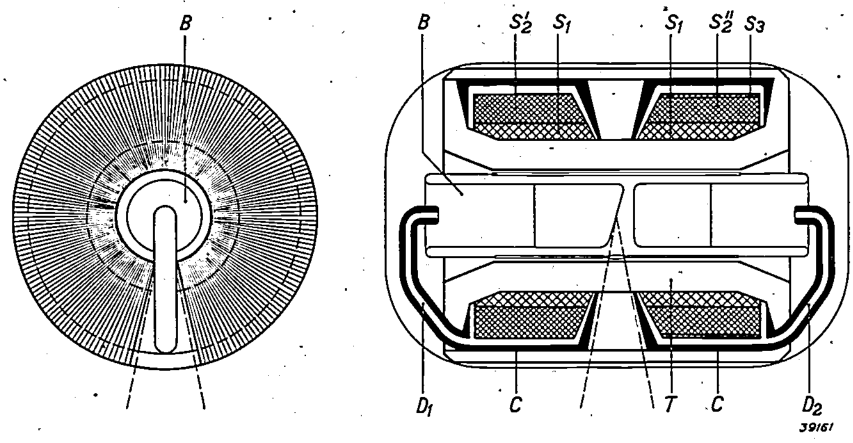


Fig. 3. Cross section of the "Centralix" apparatus. The X-ray tube B placed at the axis is drawn with fine lines. T transformer core in the form of a cylindrical hollow ring. S_1 low-voltage, S_2' and S_2'' high-voltage windings. C "Philite" cylinders, D_1 and D_2 anode and cathode leads.

The X-ray tube

The tube itself is shown in cross section in *fig. 4*. It is of the usual construction of Philips tubes in which the space between cathode and anode (discharge space) is surrounded by an earthed metal cylinder. This cylinder is here made quite long, in order to ensure adequate protection of the glass parts of the tube against secondary electrons. In order to obtain sufficiently long paths for creeping discharges between the metal cylinder and the electrodes without lengthening the tube, the glass sections of the tube which connect these parts mechanically and separate them electrically are bent as shown in fig. 4. In this way a very short

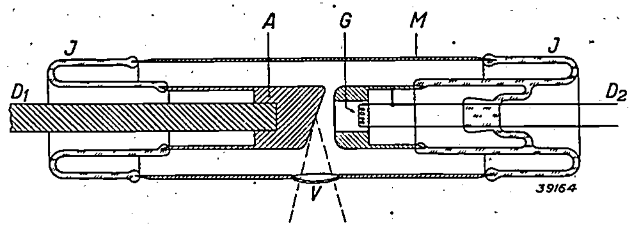


Fig. 4. Cross section of the X-ray tube of the "Centralix" apparatus. A anode, G filament surrounded by cathode cap, M earthed metal cylinder, V window for the X-rays, I bent glass insulation sections.

Fig. 5. X-ray tube of the "Centralix" apparatus, designed for a tube voltage up to about 60 kV and a tube current up to about 10 mA.

and sturdy tube is obtained, see *fig.* 5. Especial attention had to be paid to the insulation between the tube electrodes and the iron core of the transformer. In order to obtain at all points the necessary distances *via* glass and air between the earth potential plane (the iron core) and the leads at high voltage, the cylindrical opening of the transformer body has at both ends the funnel shape visible in fig. 3. Moreover, the leads are here again covered with compound.

A remarkable feature of the tube is the construction of the cathode. An ordinary X-ray tube works so far in the saturation region of the cathode emission that at 1/3 of the maximum working voltage the tube current already reaches its peak value, see *fig.* 6*a*. The result is that over a large part Δt of the period of the alternating current there is a considerable dissipation of energy, while because of the sharp increase in the efficiency of the excitation of the radiation with the voltage, it is only in the neighbourhood of the peak value of the A.C. voltage that any appreciable X-radiation is obtained. By placing a cathode cap of a certain form (a kind of "grid") around the filament, the effect of the space charge which tends to retard the emission of the filament, can be so much increased that only close to the peak value of the tube voltage does the maximum tube current begin to flow, fig. 6*b*. The ratio between the total X-ray energy and the total anode load is hereby considerably improved. This device can of course only be employed when about the same tube voltage is always used, as is the case with the "Centralix" apparatus; in the case of an ordinary tube which works with very widely diverging tube voltages, it must still work in the saturation region of the emission at the lowest voltage used, so that a situation like that in fig. 6*a* is automatically excluded at the highest voltages.

Fig. 6. *a*) Cathode emission (tube current) i as a function of the tube voltage e in a normal X-ray tube. The maximum tube voltage e_0 lies so far in the saturated region that the tube current has the saturation value over a large part Δt of the period T of the A.C. voltage supply.

b) Same as in (*a*) for the tube of the "Centralix" apparatus. The tube current now only reaches its saturation value when the A.C. voltage is near its peak value. The radiation yield is thus better than in (*a*).

Method of suspension and use of the apparatus

The tube and the transformer are surrounded by a smooth chromium-plated jacket, while the cylindrical body thus obtained, which can be rotated about its axis, is hung in a metal fork, see *fig.* 7.

Fig. 7. The "Centralix" apparatus together with relay switch mounted in a fork and provided with "Philite" cone for directing the X-ray beam.

On this fork is also the relay switch which can be set at exposure times between $^1/_4$ and 12 seconds, and which is operated with a continuous release such as is used in cameras. Over the opening through which the X-rays are emitted a "Philite" cone is screwed, by means of which it is possible to ascertain whether the central ray of the X-ray beam is directed upon the object. This device for the directing of the beam is necessary because in this case, in contrast to the case of the ordinary large X-ray apparatus, there is no fixed relative orientation of film, object and focus. The elimination of this fixed orientation was the main object in the construction of the small apparatus. Instead of the cone a lead glass or metal tube can also be used, which at the same time limits the lateral spreading of the X-ray beam, or, with greater distances between focus and film, a telescopic centring rod. This latter must of course be removed during the exposure.

Since the apparatus (without fork) with a longest dimension of 25 cm weighs only 12 kg, it is often possible to improvise some kind of support for it when it must be carried about. For use by the dentist the apparatus can be hung on a movable standard or on a wall bracket which can be extended and turned in all directions. In photographing the jaw the film is placed in the patient's mouth and pressed against the teeth to be photographed. *Fig. 8* shows a photograph which was taken this way with the "Centralix" apparatus.

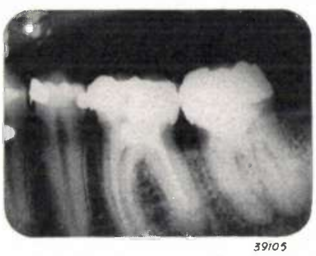


Fig. 8. X-ray photograph of teeth (back right of lower jaw), obtained with the "Centralix" apparatus. The extent and position of the metal fillings is very clearly visible.

The "Practix" apparatus

In making a compromise between required performance and permissible size of an X-ray installation, the emphasis may of course be laid as desired more on the one or on the other feature. In the case of the "Centralix" the emphasis was laid much more on obtaining a small and simple unit. Experience has, however, shown that there is also need for an apparatus where the emphasis is more in the other direction: *i.e.* an apparatus which may still be taken to the patient and adjusted, but which for the rest may be somewhat larger and more complicated in order to permit a correspondingly greater versatility. This involves not only the desirability of being able to make exposures with a somewhat larger and an adjustable current, but particularly that of being able to use the apparatus in fluoroscopy. The "Practix" apparatus meets these requirements.

The connections

This apparatus is designed for two voltages, namely one of about 60 kV which is used for photographs of the skull, shoulder, vertebrae, etc. and one of about 52 kV for photographs of the jaw, etc. [4]). The higher voltage is obtained very simply by introducing a tap on the primary winding, thus increasing the transformer ratio. In this case, however, the filament of the X-ray tube cannot be supplied directly from a winding on the high-voltage transformer, since at the lower voltage stage the filament voltage would then be 20 per cent lower and the cathode emission (the tube current) therefore about 90 per cent lower than at the higher voltage stage.

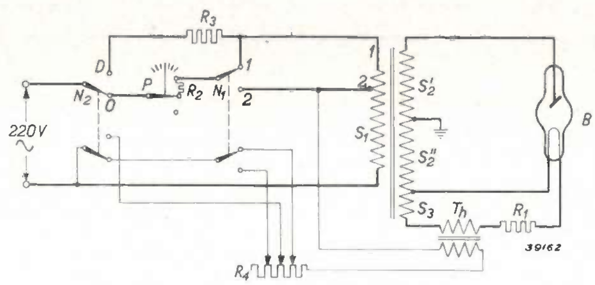


Fig. 9. Connections of the "Practix" apparatus with two voltage stages for photography. On the primary winding S_1 of the high-voltage transformer two taps *1* and *2* are introduced between which a selection is made with the switch N_1. The correct tap on the series resistance R_4 of the auxiliary transformer T_h for obtaining a tube current of 10 mA in each case is hereby also selected. With the switch N_2 one switches over to "fluoroscopy", whereby automatically a third tap on R_1 reduces the current to 3 mA. R_3 is a damping resistance, the remaining letters have the same significance as in fig. 1.

In order to avoid this, in adjustable apparatus, as was mentioned above, a separate heating current transformer is used. For the sake of saving weight and space in our case, however, a different method has been followed which is illustrated in the diagram of the connections, *fig. 9*. The filament voltage is supplied mainly by a winding S_3 on the high-voltage transformer. In series with this is a small auxiliary transformer T_h, whose primary winding is connected to the mains *via* a resistance R_4 with different taps. By connecting in series a large part of R_4 at the higher voltage, and a small part at the lower stage, the small transformer T_h supplements the filament voltage each time to exactly the desired value (tube current 10 mA). Since T_h only has the character of a correction

[4]) The lower voltage stage is omitted in the apparatus designed especially for dentists, since the higher stage is sufficient for photographs for ordinary diagnostic purposes. The commutation for fluoroscopy desribed further on, however, is retained.

element, it may have a very light construction (see below).

In fig. 9 it may be seen that the choice of low or high exposure voltage is made simply by moving the switch N_1. A second switch (N_2) serves for the transition from "photography" to "fluoroscopy". The high-voltage transformer is hereby connected to the tap (*1*) for low voltage (*via* a damping resistance R_3), while at the same time by a tap on R_4 the tube current is set at a smaller value than for photography, namely at 3 mA (since with this smaller tube current the voltage losses in the transformer, etc. also become smaller, the voltage is not about 52 kV in fluoroscopy, but about 57 kV). This decrease in tube current is possible since in fluoroscopy a lower X-ray intensity is sufficient and it is desirable since it would otherwise become much more difficult to limit the heating up of the apparatus to the necessary degree.

The construction

In order to keep the temperature of the apparatus sufficiently low in fluoroscopy, it is necessary to limit the losses in the transformer as much as possible, but at the same time to provide for a good heat dissipation from the transformer and the tube. This consideration made it necessary to give up the elegant construction principle of the "Centralix". In this case as far as the dissipation of heat is concerned, the situation is very unfavourable for the tube mounted inside the transformer as well as for the windings which are carefully insulated and shut up inside the hollow transformer, while at the same time the solid impregnating material, with which the windings as well as the ends of the tube are covered in order to limit the necessary insulation distances, may offer difficulties upon long continued heating; cavities may occur in the compound due to expansion and contraction, which cavities constitute a danger to the insulation [5]).

For these reasons the classical core transformer of fig. 2*a* was again used in the "Practix" apparatus, and oil was used as insulation material instead of compound. Thanks to the high breakdown stability (at higher temperatures also) of the oil the construction may again be made very small. While the filling with oil involves a greater weight of the apparatus and causes a structural complication in the necessity of an oil-tight seal, there is the advantage that the heat dissipation is greatly facilitated by the oil.

Fig. 10 shows a cross section of the whole apparatus. The X-ray tube, which is identical with that of the "Centralix" apparatus, is fastened against the free arm of the transformer core. A lead jacket around the tube provides the necessary protection against undesired rays. Beside the tube is the auxiliary heating current transformer (T_h in fig. 9), which could be constructed with a simple straight core (thus with an open magnetic circuit) because of the low power required, around which, separated by an insulating tube are the low and high-voltage windings. The whole is immersed in oil and surrounded by an earthed jacket which is soldered in order to provide an oil-tight seal. The filling with oil (impregnation) takes place in a vacuum in order

[5]) In the "Centralix", which is used mainly for photography, the heating effect is so small that it presents no difficulties, mainly because of the fact that the compound is only used here in thin layers in which no appreciable cavities can be formed.

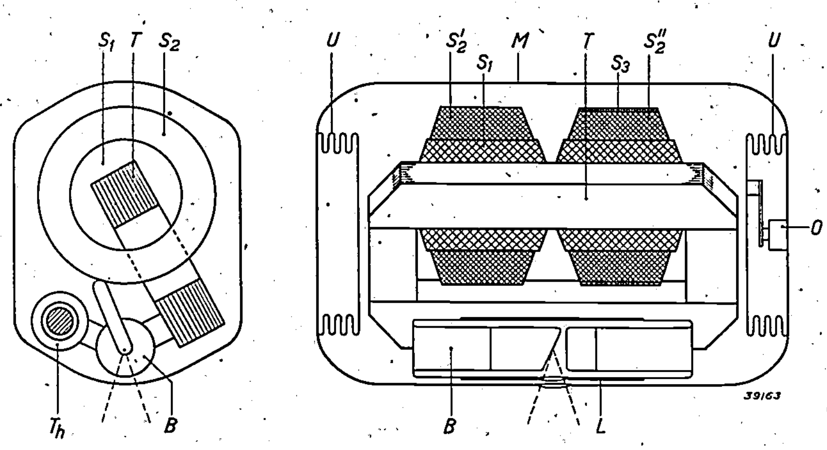


Fig. 10. Construction of the "Practix" apparatus. The high-voltage transformer has a core T of the ordinary form (square with almost square cross section) lying obliquely in the apparatus. On one arm of the core which lies about at the axis of the apparatus are the windings S_1, S_2' and S_2''; against the other arm the X-ray tube B surrounded by a lead jacket L is fastenend, and to this in turn the auxiliary heating current transformer T_h. The whole is immersed in oil and surrounded by the earthed mantle M. U are flexible cans to allow for the expansion of the oil, O switch for protection against overheating.

to avoid the presence of any residues of air and moisture which would be harmful to the insulation. Since with the temperature changes to be expected (—10° C in the winter, to +60° C after long continued fluoroscopy, for example) volume changes of about 5 per cent in the oil must be taken into account, accordion-shaped cans of very thin metal are attached to the side walls of the apparatus (fig. 10), which permit a motion of the bottoms of the cans of several millimetres and in this way the necessary expansion of the oil.

The heat balance of the apparatus

Due to the fact that the oil is set in motion by the heat during use, it may be assumed that the heat developed in tube and transformer will be distributed quite uniformly throughout the oil and thus over the whole apparatus. The power dissipated in fluoroscopy amounts to 166 watts (namely 131 watts in the tube and 35 watts in the high-voltage transformer and in the heating current circuit), while the surface of the apparatus is 1 600 cm². If we count on a heat dissipation of 10^{-3} W/cm² °C, which corresponds to that found in power oil-transformers, it will be seen that upon continuous use the temperature will rise until it is 166/1.6 = 104° above room temperature. Such a temperature increase could not of course be permitted; but actually it never reaches that point for two reasons: in fluoroscopy the tube is not actually in continual use, but there are occasional pauses, for instance for adjustment and for questioning the patient being examined and for calling the next patient. If the apparatus is switched off during these pauses which consume almost as much time as the fluoroscopy itself — and it is desirable to do so in order to expose the patient as little as possible to the X-rays — the watt consumption is reduced by one half, and the final temperature of the apparatus is limited to 20 + 104/2 = 72° C with a room temperature of 20° C. Moreover, it takes considerable time before this final temperature is reached, because of the heat capacity of the apparatus. The total heat capacity, *i.e.* for the oil and metal parts together, amounts to 2 400 cal. Beginning with the cold state (temperature T_0) the temperature of the apparatus T varies as follows:

$$T-T_0 = (T_{max}-T_0) \cdot (1-e^{-t/\Theta}),$$

where t is the time, $T_{max}-T_0$ the above calculated final temperature increase (104 or 52°) and Θ is the quotient of heat capacity and heat dissipation per second and °C. From this it may be calculated that in the intermittent use referred to (ordinary

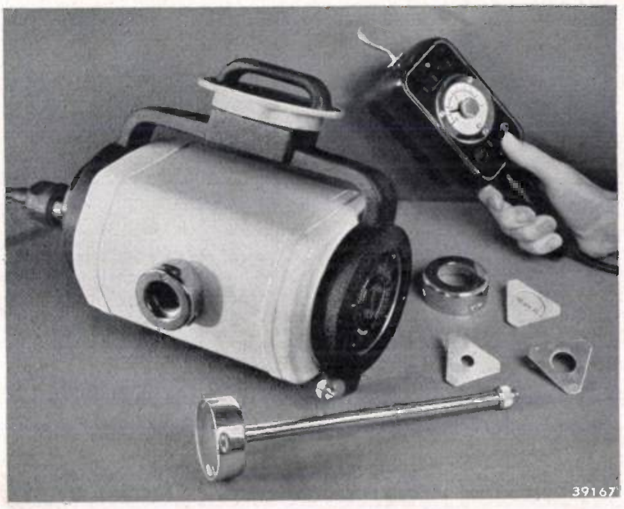


Fig. 11. The "Practix" apparatus mounted in a fork with "saucer" and handle. On the right the hand switch which contains the switches N_1 and N_2 for the two voltage stages and for fluoroscopy, as well as the timing switch and the resistance R_4. In the foreground a telescopically extensible rod which can be fastened to the window of the X-ray tube for directing the beam and several diaphragms and filters.

fluoroscopy) and with the relatively high initial temperature of 25° C, a temperature of 60° C, which is permissible, is only reached after nearly 2 hours.

It is of course necessary when, as in this case, the gain in heat is greater than the heat dissipation, to introduce a safety device to prevent overheating of the apparatus during long continued use. This could very easily be done by means of the accordion-shaped cans on the side walls mentioned above, which are compressed when the oil is heated: on the movable bottom of one of these cans there is a lever which at a given displacement, corresponding

Fig. 12. Fluoroscopy with the "Practix" without the use of a standard (examination of leg fracture).

to the highest permissible temperature operates a switch which interrupts the current supply to the high-voltage transformer. A temperature of 60° C is usually chosen as permissible limit.

The above calculation was found to be confirmed satisfactorily upon use of the apparatus in routine fluoroscopy examinations of the Medical Department of the Philips Plant. In this case for example in 1 hour and 40 minutes 70 persons were examined without the automat coming into action.

Method of support and use of the apparatus

The apparatus soldered into its housing is 30 cm long and weighs 14 kg. Like the "Centralix" it is mounted in a fork in such a way that it can rotate about its axis, see *fig. 11*. The relay switch which regulates the exposure time in the same way as in the "Centralix" is housed with the two switches N_1 and N_2 in a so-called hand switch. In this hand switch is also the regulation resistance R_4 (see fig. 9), with which, without opening the apparatus proper, the tube currents for photography and fluoroscopy can be further regulated if necessary.

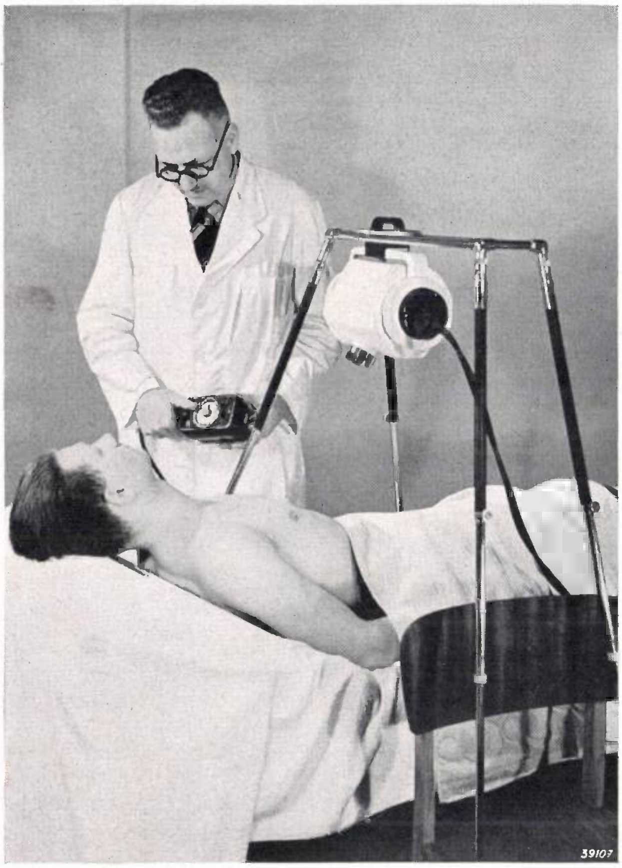


Fig. 13. Method of supporting the "Practix" with its saucer in the groove of a portable folding four-legged standard (lung photograph in patient's own home; lung photographs of satisfactory quality can be made with this apparatus).

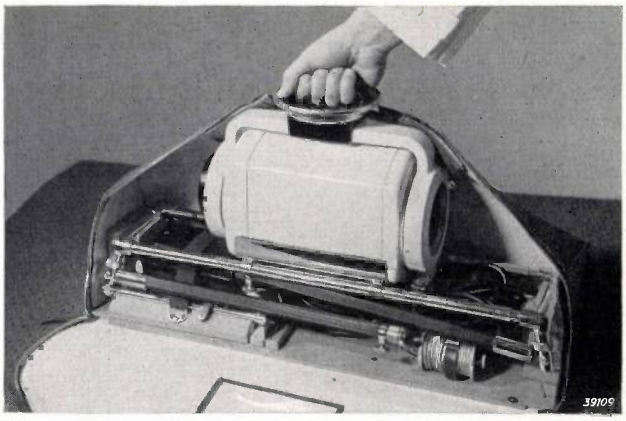


Fig. 14. The "Practix" apparatus, together with the folding four-legged standard and other accessories are housed in a small case.

Although it is also possible to take photographs with the "Practix" with improvised supports and also to use it for fluoroscopy in this way (*fig. 12*), for most cases it is nevertheless desirable to be able to fasten the apparatus to a standard. Great care was devoted to satisfying this desire without sacrificing the primary requirement of portability and adjustability in the case of a patient who cannot be moved. The simplest standard, which can easily be carried, is a folding four-legged one with extensible legs (*fig. 13*). By means of the ring with flange under the handle of the apparatus (so-called "saucer", see fig. 11) the apparatus can be hung between the two parellel horizontal tubes of the standard and can then be rotated not only about its own horizontal axis but also about a vertical axis. When folded, the standard with the apparatus and appurtenances can be housed in a

Fig. 15. Control photograph of a so-called nailing of the neck of the femur taken with the "Practix" apparatus during the operation. The photograph shows that a good picture can also be obtained of such a relatively difficult object.

small case (*fig. 14*); the total weight of this portable installation is only 25 kg, so that, for example, the general practitioner can easily take it with him in his car when he visits a patient in the country.

In addition a somewhat heavier and more convenient standard on wheels has also been developed, with which the apparatus can be fixed at any desired height between about 30 and 170 cm above the ground and at 20 to 60 cm distance from the column of the standard, while the X-ray beam can still be directed in any desired direction. Combined with this standard the "Practix" apparatus is particularly well suited for hospitals and field dressing stations. It has been in such use for a year with good results for the examination of wounded, the checking of the correct position of plaster casts and extensions, the control at the operating table of difficult surgical manipulations (*fig. 15*) and similar applications. It is the portability and ease in adjusting the apparatus which are here found to lead to a wider application of X-ray examination than would be possible if only large apparatus are used.

AN APPARATUS FOR THE INVESTIGATION OF SHARPNESS OF HEARING (AUDIOMETER)

by L. BLOK and H. J. KÖSTER.

534.771

Ear specialists are beginning to make more and more use of electrical apparatus which give continuous pure tones of variable intensity instead of the mechanical-acoustical devices which were formerly used in testing the hearing. Such an "audiometer", which has been developed from the tone generator GM 2 037 described previously, is here discussed. The method of employment in the determination of the sometimes very low thresholds of auditory sensitivity makes very strict requirements as to the freedom from interference of the signal of the audiometer. The measures are described by which these requirements can be satisfied. A discussion is also given of a number of auxiliary arrangements needed by the ear specialist in his examination. The results of the examination can be recorded in the form of an "audiogram" which gives the hearing loss in decibels as function of the frequency, and which may serve as one of the bases for the diagnosis and subsequent treatment.

The deviations in the sense of hearing observed in the case of persons who are partially deaf may exhibit great individual differences quantitatively as well as qualitatively. Quantitatively in the degree of depreciation of the sensitivity of the ear compared with that of a normal person, and qualitatively in the manner in which the deviation is manifested for the different pitches. Cases are known for example in which it is mainly the high tones which are heard poorly, and others in which the sensitivity is decreased only for the low tones.

The object of the ear specialist who desires to make an accurate diagnosis of a given case will in the first place be to measure the depreciation in the sharpness of hearing as a function of pitch. In addition he will attempt to determine the part of the ear in which the deviation is localized. The information obtained will not only furnish guidance for the therapeutic treatment to be applied, but will at the same time be important in deciding whether the patient should use some kind of hearing instrument and of what type this should be.

Until now the ear specialist has generally used a set of tuning forks to measure the decrease in ear sensitivity for different pitches. These tuning forks were struck in turn in a given way and allowed to vibrate freely. The patient then had to state for each tuning fork at what moment he no longer heard the sound. It is obvious that with a given initial intensity and a given damping (decrement) of the sound of the tuning fork, the patient will cease to hear the sound sooner, the lower his sensitivity for that pitch. The degree of depreciation of the ear sensitivity compared with that of a normal person was then indicated directly by the specialists by the "shortening" in seconds.

The objections to this very simple method are immediately obvious. The damping of a tuning fork depends very closely upon its frequency as well as upon all kinds of details of its construction. On the one hand, therefore, the results of different examiners using slightly different tuning forks may vary considerably. On the other hand the "shortening" found is not a good measure of the imporsmall case (*fig. 14*); the total weight of this portable installation is only 25 kg, so that, for example, the general practitioner can easily take it with him in his car when he visits a patient in the country.

In addition a somewhat heavier and more convenient standard on wheels has also been developed, with which the apparatus can be fixed at any desired height between about 30 and 170 cm above the ground and at 20 to 60 cm distance from the column of the standard, while the X-ray beam can still be directed in any desired direction. Combined with this standard the "Practix" apparatus is particularly well suited for hospitals and field dressing stations. It has been in such use for a year with good results for the examination of wounded, the checking of the correct position of plaster casts and extensions, the control at the operating table of difficult surgical manipulations (*fig. 15*) and similar applications. It is the portability and ease in adjusting the apparatus which are here found to lead to a wider application of X-ray examination than would be possible if only large apparatus are used.

AN APPARATUS FOR THE INVESTIGATION OF SHARPNESS OF HEARING (AUDIOMETER)

by L. BLOK and H. J. KÖSTER. 534.771

Ear specialists are beginning to make more and more use of electrical apparatus which give continuous pure tones of variable intensity instead of the mechanical-acoustical devices which were formerly used in testing the hearing. Such an "audiometer", which has been developed from the tone generator GM 2 037 described previously, is here discussed. The method of employment in the determination of the sometimes very low thresholds of auditory sensitivity makes very strict requirements as to the freedom from interference of the signal of the audiometer. The measures are described by which these requirements can be satisfied. A discussion is also given of a number of auxiliary arrangements needed by the ear specialist in his examination. The results of the examination can be recorded in the form of an "audiogram" which gives the hearing loss in decibels as function of the frequency, and which may serve as one of the bases for the diagnosis and subsequent treatment.

The deviations in the sense of hearing observed in the case of persons who are partially deaf may exhibit great individual differences quantitatively as well as qualitatively. Quantitatively in the degree of depreciation of the sensitivity of the ear compared with that of a normal person, and qualitatively in the manner in which the deviation is manifested for the different pitches. Cases are known for example in which it is mainly the high tones which are heard poorly, and others in which the sensitivity is decreased only for the low tones.

The object of the ear specialist who desires to make an accurate diagnosis of a given case will in the first place be to measure the depreciation in the sharpness of hearing as a function of pitch. In addition he will attempt to determine the part of the ear in which the deviation is localized. The information obtained will not only furnish guidance for the therapeutic treatment to be applied, but will at the same time be important in deciding whether the patient should use some kind of hearing instrument and of what type this should be.

Until now the ear specialist has generally used a set of tuning forks to measure the decrease in ear sensitivity for different pitches. These tuning forks were struck in turn in a given way and allowed to vibrate freely. The patient then had to state for each tuning fork at what moment he no longer heard the sound. It is obvious that with a given initial intensity and a given damping (decrement) of the sound of the tuning fork, the patient will cease to hear the sound sooner, the lower his sensitivity for that pitch. The degree of depreciation of the ear sensitivity compared with that of a normal person was then indicated directly by the specialists by the "shortening" in seconds.

The objections to this very simple method are immediately obvious. The damping of a tuning fork depends very closely upon its frequency as well as upon all kinds of details of its construction. On the one hand, therefore, the results of different examiners using slightly different tuning forks may vary considerably. On the other hand the "shortening" found is not a good measure of the importance of the deviation. If, for example, two frequencies are considered for which the free vibration of the corresponding tuning forks is similar to that shown in *fig. 1a* and *b*, it is seen that a given depre-

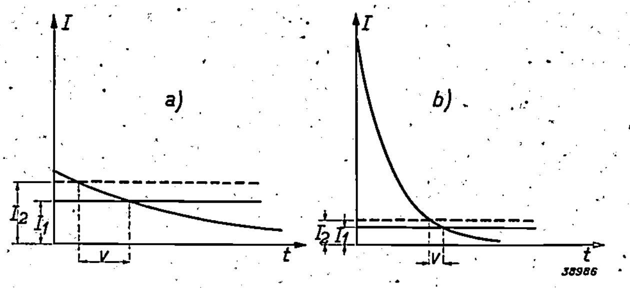


Fig. 1. Variations of the sound intensity upon the free vibration of a tuning fork of low frequency (a) and of high frequency (b). In both cases the normal threshold of hearing for the frequency in question (I_1) is represented by a full line, while the dotted line represents the raised threshold of hearing of a person with defective hearing (I_2). With a given ratio I_2/I_1, *i.e.* with a given loss of hearing in decibels, a quite different "shortening" v is found in the two cases.

ciation in ear sensitivity (*i.e.* a given increase in the threshold value measured in decibels) gives quite different "shortenings" in the two cases [1]). Moreover, the sound intensity obtainable with tuning forks is often so small for low frequencies (fig. 1a) and the threshold value of the ear for these frequencies so high, that even with only relatively slightly deficient hearing the patient hears nothing at all from the very beginning, and thus no "shortening" can be measured at all, while with tuning forks for high frequencies (fig. 1b) the damping is often so great that it is difficult for the patient to indicate the exact moment at which for him the sound ceases, so that the measurement becomes very inexact.

These disadvantages have led to an attempt which has become more definite in recent years to replace the tuning forks and the other mechanical acoustical devices of the ear specialist by more modern apparatus which produces instead of damped tones continuous tones of variable intensity.

The obvious solution of this problem is to use the oscillators which are employed technically in the testing of parts of electro-acoustic installations, such as amplifiers, cables, loud speakers, etc. These oscillators give sinusoidal A.C. voltages with frequencies which can be varied over the whole acoustic range and with adjustable intensity. Such an apparatus, usually called a "tone generator", although for the original purposes it did not need to produce any "tone", has been described earlier in this periodical [2]). By connection to a loud speaker and several alterations in the construction, as well as by the addition of various auxiliary arrangements, an apparatus could be constructed from this tone generator which is especially adapted for the testing of hearing. The apparatus obtained (audiometer) and its use will be discussed briefly in the following.

Construction of the audiometer

Fig. 2 gives a very much simplified diagram of the original tone generator (type GM 2 307). For practical reasons [3]) the A.C. voltage of the desired

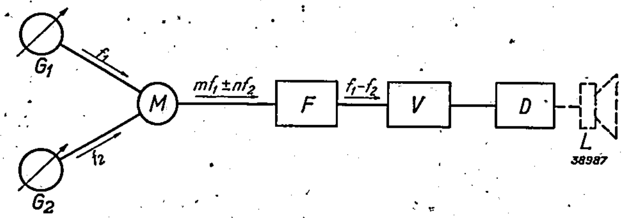


Fig. 2. Diagram showing the principle of the tone generator GM 2 307. G_1 and G_2 oscillators, M mixing valve, F filter, V amplifier, D attenuator. A loud speaker L may be connected behind D.

low frequency is not generated directly by an oscillator, but is obtained as the beat between two higher frequencies f_1 and f_2. The two oscillators G_1 and G_2 generate voltages of these two high frequencies; the voltages are fed to the mixing valve M, in the anode current of which, therefore, all combination frequencies $mf_1 \pm nf_2$ occur. The signal with the desired low frequency f_1—f_2 is filtered out by the filter F and raised to the necessary level in the amplifier V (normally 200 milliwatts). The output frequencies f_1 and f_2 of the oscillators G_1 and G_2 can be varied by means of rotating condensers between 100 and 101 or 100 and 85 Kc/s. By this means the differential frequency f_1—f_2 obtained can be varied from 0 to 16 kc/s, *i.e.* within the whole acoustic frequency range. The regulation of the intensity is by means of the adjustable attenuator D.

For testing the hearing a loud speaker is now connected behind the attenuator and the patient listens to the sound from the loud speaker. With the help of the attenuator the doctor can decrease the intensity of the sound to such a degree that the patient just barely hears it. If the attenuation factor is d_1 decibels in this case, a greater attenuation d_2

---

1) The number of decibels hearing loss is also not an entirely true measure of the importance of the deviation when different frequency regions are compared. It is, however, a much more useful standard than the "shortening" and has, moreover, the advantage of being exact and reproducible.

2) L. Blok, A Tone Generator, Philips techn. Rev. 5, 263, 1940.

3) See the article referred to in footnote 2).

will then be necessary for a normal person to reach the threshold of ear sensitivity in a similar way. The difference d_2—d_1 is the loss of hearing in decibels (depreciation in sensitivity of the ear) of the person with deficient hearing for the frequency in question.

It is clear that at least for the calibration of the audiometer it is also necessary to attenuate the tone produced by the instrument to the threshold of ear sensitivity for a normal person. Since this threshold is extremely low — at 1 000 c/s the normal ear perceives a sound intensity of 10^{-16} watts/cm^2 — much greater care must be taken in the case of the audiometer than in that of the ordinary applications of the tone generator that there are no disturbing sounds. The hum of the alternating-current supply (the 50 c/s of the mains) must be considered as such, as well as the noise present in every amplifier, the hum of the output transformer, the overtones caused by non-linear distortion and finally the surges occurring when a tone is switched on.

The amplitude of the hum voltage in the tone generator amounted to from $^1/_2$ to 1 per cent of that of the effective signal. This means that the ear of the patient, in addition to the desired tone, would also always be exposed to a tone of 50 c/s with a 40 decibel lower intensity. If we consider the ear-sensitivity curve of a normal person, see *fig. 3*, it

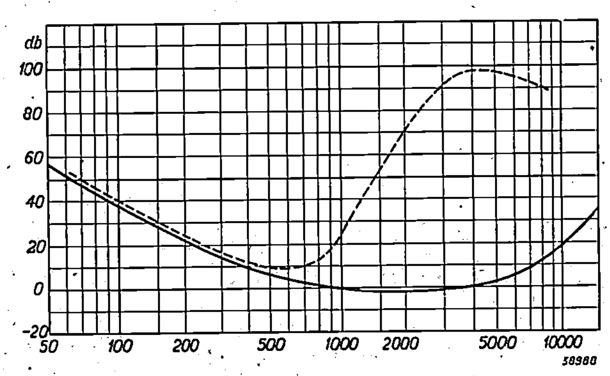


Fig. 3. Variation of the threshold value of sound intensity for a normal ear as a function of the frequency (full line curve, according to S. S. Stevens and H. Davis, Hearing, Wiley and Sons, New York 1938, p. 50). The threshold for a frequency of 1 000 c/s is arbitrarily set equal to 0 decibels. In a certain case of deficient hearing the broken-line curve may for example be found.

is clear that this cannot cause any error, since in the case of no frequency does the threshold lie 40 decibels higher than the threshold for 50 c/sec, on the contrary, it lies much lower for most frequencies. Let us now suppose, however, that a patient must be examined whose ear sensitivity is very much decreased only for high tones, for instance according to the broken line threshold curve in fig. 3. At a frequency of 4 000 c/s, for instance, this patient would not yet be able to hear the signal itself, but would, however, already be able to hear the 40 decibel weaker hum tone for 50 c/s, so that a false picture of the ear sensitivity would be obtained. In order to decrease the chance of such errors as much as possible special measures have been taken to make the hum of the apparatus still weaker. The plate voltage of the various amplifier valves is still more carefully smoothed than in the tone generator, and although the cathodes of the valves are of the indirectly heated type which of itself gives little cause for hum, the cathodes are nevertheless fed with direct current. The required D.C. voltage is furnished by a small selenium rectifier with a simple smoothing filter. In this way the hum is so limited that even with a loss of hearing of 105 dB at 1 000 c/s, which must be considered as total deafness, no errors can occur.

Similar considerations hold for the noise. The level of the noise, however, in the tone generator GM 2 307 is already so low that no special measures needed to be taken to combat it.

In the case of the hum of the output transformer of the amplifier V (see fig. 2), as a result of the mechanical forces of the magnetic field a part of the output is converted directly into acoustic energy, thus without passing through the attenuator and loud speaker. By keeping the patient sufficiently far away from the audiometer this source of interference could of course be eliminated immediately. In practice, however, the doctor who operates the audiometer and the patient will usually be seated at a distance of only a few metres from each other, and at a distance of 3 m for example the hum of the transformer was still observable in a quiet room. From the nature of the case the effect can only lead to errors when the signal of the loud speaker is very weak, *i.e.* when the attenuation factor is very large. This hum was therefore rendered harmless in the following simple manner. The attenuator, which permits a maximum attenuation of 155 dB — this is more than sufficient since with an attenuation of about 105 dB the lowest threshold which occurs is already reached —consists of a variable part with eleven steps of 5 dB each and two fixed parts each with an attenuation of 50 dB. The latter are only switched on when the eleven steps of the variable part have been traversed, thus when attenuations greater than 40 dB have been reached. Of these fixed attenuators, which are built up of a number of damping cells, a part with an attenuation of 25 dB is now, however, not connected behind, but in front of the amplifier V, see *fig. 4*. The voltage on the input of the amplifier and therefore also that on the output transformer is thus now also attenuated by this factor. By this means the hum is made inobservable even with the highest attenuation factors.

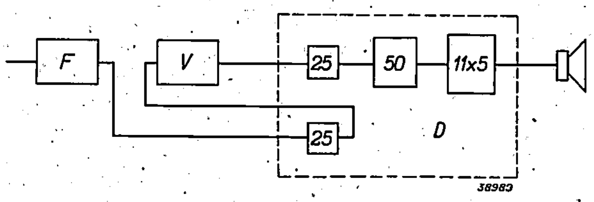


Fig. 4. The attenuator D consists of a part which can be regulated in 11 steps of 5 dB each and two fixed parts each of 50 dB which can be put into connection in turn. The first of these two 50 dB elements is split into two parts of 25 dB, one of which is connected in front instead of behind the amplifier V.

At the same time this measure is useful for combatting the errors due to overtones. These errors may occur especially with low frequencies, not only because the distortion of amplifiers is in general greatest here, but chiefly because of the characteristic shape of the normal ear sensitivity curve in this region. If for example it is desired to measure at a frequency of 50 c/s, the second harmonic of the signal (100 c/s) must according to fig. 3 be at least 18 dB weaker than the fundamental in order that the normal ear shall not hear the second harmonic rather than the fundamental tone. With deviating ear sensitivity curves still steeper slopes (*i.e.* still greater threshold differences within an octave) may occur, so that still higher requirements are then made of the freedom from distortion. In the case of the tone generator GM 2 307 the distortion amounts to an average of $^1/_2$ per cent with the maximum output (at low frequencies the maximum is 1 per cent), so that threshold differences of 40 dB in an octave can still be measured satisfactorily. With not too much decreased ear sensitivity, however, the performance of the audiometer is still considerably better, since then, in the manner described, an attenuating element of 25d B is connected in front of the amplifier, so that it has a much weaker signal to deal with and therefore much less non-linear distortion occurs.

The connection of a part of the attenuator in front of the amplifier is, however, a disadvantage for the level of noise. The noise is to a large extent generated by the first amplifier stage and will be of less importance relatively, the stronger the input signal of the amplifier. It was for this reason that the whole attenuator was not placed in front of the amplifier, which would of itself have been simpler and better for the desired combatting of the transformer hum and the distortion. In this way the noise is limited to such a low level that it can only become audible at a sound intensity of 70 dB above the threshold at 1 000 c/s.

Finally there remain the switching surges. When the loud speaker is switched on in the ordinary way there is always a click which may be considerably louder than the tone to which the patient must listen. This must be avoided. For this purpose the simple connections of the oscillator G_1 (fig. 2) shown in *fig. 5* are employed. As long as no tone is desired the switch S is closed and the condenser C is charged to a voltage of 20 volts. On the control grid of the oscillator valve therefore there is a negative D.C. voltage of 20 volts, the valve is overbiased and cannot oscillate. In order to switch on the tone the switch S is now opened. The condenser is slowly discharged *via* the large resistance R, the grid bias of the valve becomes gradually less negative and the valve begins to oscillate with an amplitude which grows slowly to the final value. Upon switching on the tone therefore no click occurs and the sound gradually (within 1 sec for instance) takes on the previously chosen intensity.

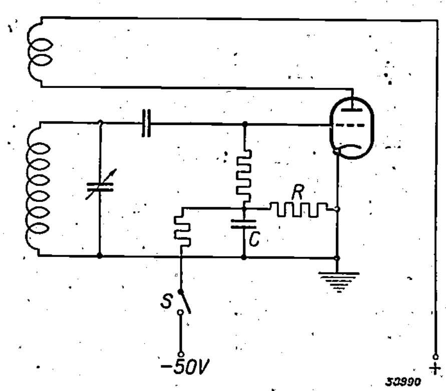


Fig. 5. Connections of the oscillator G_1 with RC circuit for the gradual raising of the oscillations to their final amplitude (avoidance of clicking sounds in the loud speaker). With the switch S closed the condenser C is charged to a voltage of 20 volts.

The complete arrangement

Fig. 6 is a photograph of the audiometer with several auxiliary instruments which are used in the examination. On the audiometer itself may be seen the dials of the two rotating condensers with which the desired pitch is obtained. The frequency of the tone produced is given simply by the sum of of the frequencies read off on the two dials. Above may be seen the knobs of the attenuator, a measuring instrument for insuring that the signal voltage without attenuation has the normal initial value, and below various switches and lamps for signaling purposes.

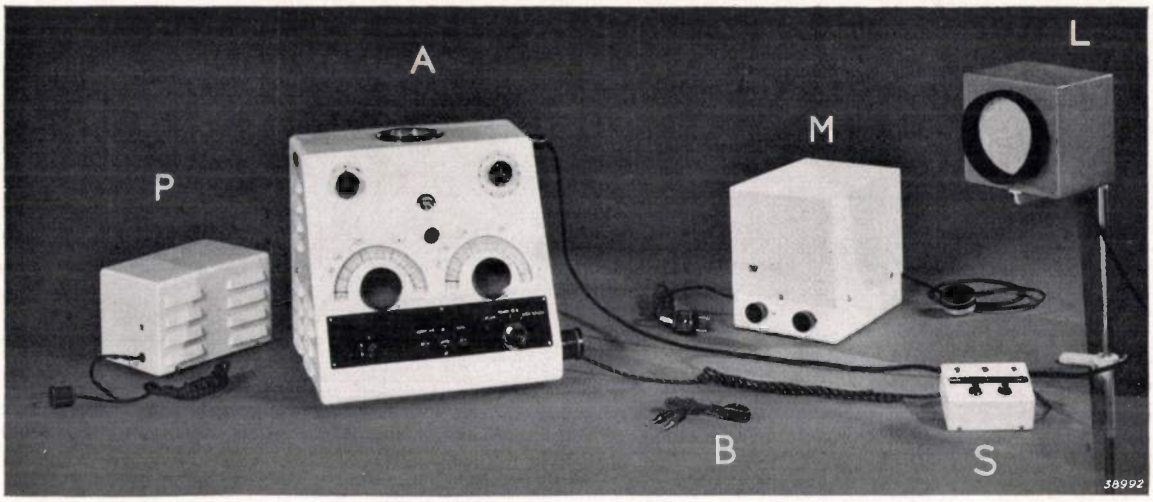


Fig. 6. Complete arrangement for the testing of the ear. *A* audiometer with frequency scales, attenuator knobs, etc. *P* necessary supply apparatus. *L* loud speaker, *S* box which is used by the patient for signalling. The loud speaker may if desired be replaced by the bone conduction telephone *B*. *M* auxiliary apparatus for so-called masking.

For the examination to take place smoothly the necessary communication between doctor and patient is by means of light signals. Both doctor and patient have three lamps in front of them, yellow, green and red (those of the patient are on the small box below the loud speaker in fig. 6). When the doctor switches on the tone the yellow lamp lights up in both cases. The patient must now press one of two buttons according to whether he hears the sound or not. If he presses the button "yes" the green lamps light, if the button "no" the red lights. In this way the threshold of hearing is very quickly and rapidly determined. By means of a switch the doctor can also make the tone continuous and then by variation of the frequency he can quickly determine the limits of hearing as far as pitch is concerned. Finally the doctor also has at his disposal a "simulator's switch". With this switch the tone generator is put out of action, while the signal arrangement continues to function. Then upon the lighting of the yellow lamp the patient should always answer "no", if he answers "yes" he is evidebtly desirous of making his hearing appear better than it is. Fatigue of the patient which no longer allows him to distinguish with confidence whether he hears the sound or not can also be detected by the doctor in this way.

A very small type of loud speaker has been chosen. This is desirable since the two ears of the patient must be examined separately, and with a large loud speaker it is difficult to restrict the influence of the second ear sufficiently. With the small loud speaker used this is provided for by a cylinder of felt against which the patient presses his head. The shielding of the free ear would be still better with the use of a head phone. This has not been employed, however, because in the first place resonance may occur in the relatively small column of air between membrane and ear drum which would cause the delusion of an increased sensitivity for some frequencies; and in the second place because the housing of the telephone pressed against the external ear also vibrates to a certain extent and may therefore stimulate an impression of sound through the bony structure.

This leads us to another important point in the ear examination: the localization of the defect. A depreciation in the sensitivity of the ear may be caused by defects in the inner ear (labyrinth) or by defects in the external and middle ear which conducts the sound to the inner ear (meatus, ear drum, bones of the ear). In the first case, which may for instance occur as a result of some disease or other, like scarlet fever, one speaks of perception deafness, in the second case of conduction deafness. The method of distinguishing between the two cases is by the bone conduction just mentioned: if the middle ear is defective the patient will still be able to obtain a satisfactory sound impression *via* the bone conductors; of the defect is in the inner ear the bone conduction will be of as little use as the air conduction *via* the ear drum. For this examination a bone conduction telephone (see fig. 6) has been provided with the audiometer. This is essentially an ordinary telephone, but one which is so constructed that the moving membrane makes only small deviations but can thereby exert high pressures. The membrane (actually the membrane is formed by the whole housing of the telephone) is pressed against the skull behind the ear. The ear-sensitivity curve of a normal person has of course a different form for bone conduction than the curve of fig. 3. We shall not, however, go into that here.

The necessity already mentioned of only allowing sound to reach the ear which is being tested presents another problem to the examiner when he encounters a case of so-called unilateral deafness, in which one ear functions normally or approximately so. If in this case the defective ear has a considerable hearing loss, so that high sound intensities are necessary to reach the threshold of hearing, it is often unavoidable that the normal ear already observes a sound *via* air conduction around the head or *via* bone conduction through the head, while the defective ear still hears nothing. In order to eliminate this disturbing effect, or at least to remove its effect on the results of the measurement the normal ear is "masked". The normal ear is exposed to a "noise" of such an intensity that the ear takes on a sort of artificial deafness for the tone to be observed, while the perception of the defective ear is practically unafffected. As masking noise a tone may be used which is very rich in overtones and thus has a very sharp sound. The fundamental is best chosen in the same frequency region as the pure tone to which the defective ear is exposed. A small auxiliary apparatus has been constructed for obtaining the masking; it is shown in fig. 6 to the right of the audiometer. It is essentially a small tone generator giving a very distorted signal which can be set by means of a switch on five different fundamental frequencies appropriately distributed over the whole acoustic range.

Instead of the examination with pure tones it may sometimes also be desirable to carry out direct intelligibility tests, *i.e.* to discover the intensity level at which ordinary speech is understood by the patient. For this purpose the audiometer is provided with a connection for a gramophone pick-up, with which speech recorded on gramophone plates can be supplied to the amplifier input.

In conclusion a few words must be said about the working out of the results obtained. They can conveniently be recorded in an "audiogram", like that shown in *fig. 7*. The loss of hearing in decibels, *i.e.* the increase in the threshold value compared with that of a normal person, is here plotted as a function of the frequency, and the measured points are the successive octaves of the note C (64 c/s) according to the custom of ear specialists. The example reproduced is a typical case of unilateral deafness. The left ear is practically normal (broken line curve); for the right ear the full line curve is measured without masking, but the maximum in the neighbourhood of 1 000 c/s, which corresponds to a small maximum in the broken line curve, immediately suggests that in this threshold curve the normal ear also collaborated. With masking the much lower dot-dash curve was indeed found, which indicates a practically total deafness of the right ear.

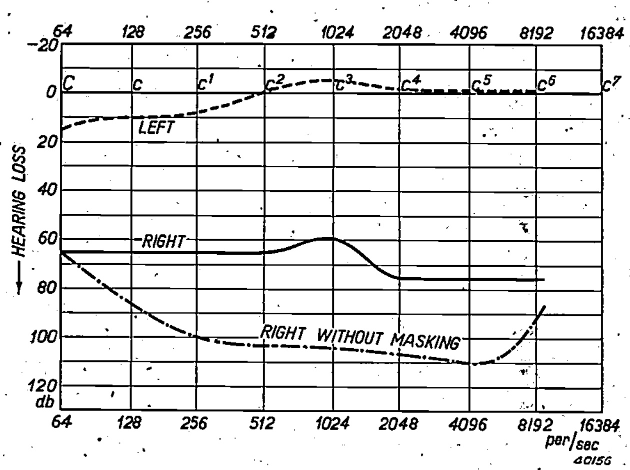


Fig. 7. Example of an audiogram. The hearing loss in dB is plotted directly as a function of the frequency. This was a typical case of unilateral deafness.

The audiogram now forms one of the bases of the diagnosis to be made and serves at the same time as documentation for later control of whether the defect has developed further and in what direction this has taken place. Finally the audiogram also furnishes the instrumentmaker with the necessary data when the ear specialist wishes to prescribe the use of a sound amplifier with an especially adapted frequency characteristic [4]).

[4]) See for example K. de Boer and R. Vermeulen, On improving of defective hearing, Philips techn. Rev., 4, 316, 1939.

LECHER SYSTEMS

by C. G. A. von LINDERN and G. de VRIES. 538.566.5

The behaviour of Lecher systems with respect to travelling and stationary waves is discussed. The way is discussed in which Lecher systems can be employed as supply connections in transmitters, as high impedances in electrical circuits, as stabling resonators for high frequencies, etc.

Two electrical conductors which are strung parallel to each other, or which in certain cases are concentric, together form a Lecher system. In the last decade of the previous century many experiments had already been performed with such systems and J. J. Thomson, among others, pointed out their great significance in the excitation and propagation of electromagnetic waves. They were later used on a large scale in radio technology. As was already mentioned in an article in the previous number of the periodical [1]), the behaviour of such Lecher systems can be described to an certain extent by means of quasi-stationary concepts, although they are not fundamentally quasi-stationary systems. We shall go into this in somewhat more detail.

If we apply to the input of a Lecher system (left in *fig. 1*) a voltage V_1 of such a high frequency

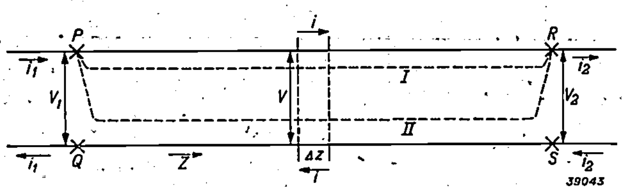


Fig. 1. Lecher system with input PQ and output RS, to which the voltages V_1 and V_2, respectively, are applied, and in which the currents i_1 and i_2, respectively, flow.

that the wave length is still large compared with the distance between the wires but not large compared with the length of the Lecher system, the currents will indeed be equal and opposite in every vertical cross section of the two conductors, but in the direction of length of the conductors the current i changes from point to point, as indeed does the voltage V between the two conductors. The change in current occurring along the conductors is a result of the displacement currents which flow through the capacity which exists between two conductors. Since the conductors have the same thickness at all points and are everywhere the same distance apart, it is possible to ascribe to a Lecher system a capacity C^I per unit of length.

In addition to this capacity C^I per cm, a Lecher system also has a self-induction L^I per cm; this is by definition the magnetic flux enclosed per cm length of the Lecher system at a current of 1 ampere. It is then found possible with the help of this self-induction L^I and capacity C^I uniformly distributed along the conductor to describe the behaviour of a Lecher system satisfactorily, and actually to apply to it the equivalent circuit of *fig. 2*. Even though the length of the Lecher

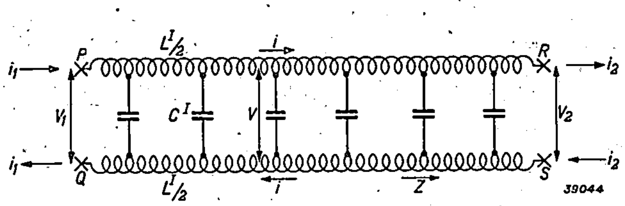


Fig. 2. Equivalent circuit for the Lecher system of fig. 1, which has a self-induction L^I per cm and a capacity C^I per cm.

wires is not small compared with the wave lengths, the Lecher system can, nevertheless, be treated in its transverse dimensions as quasi-stationary, and we shall call such a case semi-quasi-stationary. In our semi-quasi-stationary theory the difference in voltage between the points P and R or Q and S in fig. 1 may then not occur, since they are not defined [2]), but this is no disadvantage since it is only the differences in voltage between P and Q or R and S which are important in the theory.

Travelling waves

It is a well known phenomenon that waves travelling along a Lecher system can be propagated with a very definite velocity, and we shall now in the first place treat this phenomenon in a simple way [3]). For this purpose we consider a charge e (positive on one wire, negative on the other) distributed evenly over a length l of the Lecher wires, which causes a voltage V between the two

[1]) Resonance circuits for very high frequencies, Philips techn. Rev. 6, 217, 1941.

[2]) It makes a considerable difference with regard to the necessary energy whether in fig. 1 a unit charge is moved from P to R along curve I or along curve II.

[3]) *Cf.* H. G. Möller, Grundlagen und mathematische Hilfsmittel der Hochfrequenztechnik (Springer, Berlin 1940) p. 172. See also O. Heaviside, Electromagnetic theory, Vol. III, p. 3 (1893, republished in 1922, Benn Brothers, London).

wires (*fig. 3*). The following then holds:

$$e = C^I l V. \quad (1)$$

If these two charges which are originally distributed between $x = 0$ and $x = l$ move with linear velocity v towards the left, the current is :

$$i = C^I \, V \, v \quad (2)$$

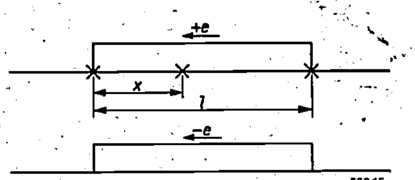


Fig. 3. A charge $+e$ is uniformly distributed over a length l of one of the Lecher wires; on the other wire in the same way is the charge $-e$.

The magnetic flux Φ between 0 and l at first amounts of $\Phi = l L^I i$ and decreases with time as the charge in its movement toward the left moves out of the region between $x = 0$ and $x = l$. The variation of Φ with the time is given by

$$\Phi = (l - vt) \, L^I i = (l - vt) \, L^I C^I \, V v \quad (3)$$

The voltage between the two conductors at the point x is according to the law of induction equal to the decrease in Φ per unit of time:

$$V = -\frac{\mathrm{d}\Phi}{\mathrm{d}t} = L^I \, C^I \, V v^2 \quad (4)$$

From (4) the square of the velocity is obtained:

$$v^2 = \frac{1}{L^I C^I} \quad (5)$$

Any given distribution of charge will also be propagated at this velocity since such a distribution can always be built up of different charge distributions according to fig. 3. This simple result also follows directly from the electromagnetic vibration equations, which we shall now set up for a short section Δz of a Lecher system, which we may consider as a current circuit with self-induction $L^I \Delta z$ and capacity $C^I \Delta z$, while we neglect the ohmic resistance. The increase Δi of the current over a length Δz of the Lecher system is then given by

$$\Delta i = -\frac{\mathrm{d}}{\mathrm{d}t} \, (V \cdot C^I) \, \Delta z,$$

so that we obtain the following partial differential equation:

$$\frac{\partial i}{\partial z} = -C^I \frac{\partial V}{\partial t} \quad (6)$$

According to (3) and (4) the voltage difference ΔV excited over a length Δz of the conductors is in the same way equal to

$$\Delta V = -\frac{\mathrm{d}}{\mathrm{d}t} \, (i \cdot L^I) \Delta z,$$

which gives us the partial differential equation

$$\frac{\partial V}{\partial z} = -L^I \frac{\partial i}{\partial t} \quad (7)$$

By partial differentiation of these two electromagnetic equations (6) and (7) with respect to z and t we find the same equation for V and i:

$$\left.\begin{aligned} \frac{\partial^2 V}{\partial z^2} &= + L^I C^I \frac{\partial^2 V}{\partial t^2} \\ \frac{\partial^2 i}{\partial z^2} &= + L^I C^I \frac{\partial^2 i}{\partial t^2} \end{aligned}\right\} \quad (8)$$

These formulae are analogous to the familiar equations for mechanical waves which are propagated along a stretched string, where instead of V or i we are concerned with the displacement of a point on the string, while instead of L^I and C^I the mass per unit of length and the reactionary force occur. Every twice differentiable function of $t \pm z\sqrt{L^I C^I}$, the so-called d'Alembert solution, satisfies equation (8). This means that every disturbance of the equilibrium which acts on the system at a given moment furnishes a solution to the equation if it is propagated unaltered in the negative or positive z direction with a velocity $v = 1/\sqrt{L^I C^I}$, as was the case with our rectangular charge distribution at the beginning of this section.

If we now take for the voltage a wave moving to the right with the velocity $1/\sqrt{L^I C^I}$:

$$V = f\,(t - z\sqrt{L^I \, C^I}), \quad (9)$$

then with the help of (6) and (7) we calculate for the current

$$i = +\sqrt{\frac{L^I}{C^I}} \, f\,(t - z\,\sqrt{L^I \, C^I}) \quad (10)$$

This is thus exactly the same form as for V, while the phases of i and V also correspond exactly at any point along the Lecher system. The quotient of the momentary values of voltage and current is constant, and the following is valid at every point along the Lecher system:

$$\frac{V}{i} = \sqrt{\frac{L^I}{C^I}} = \zeta. \quad (11)$$

ζ is called the wave resistance of the Lecher system.

This behaviour of the Lecher system is quite analogous to what occurs in an aerial which is emitting electro-magnetic waves. The energy which is thereby radiated is manifested as an apparent resistance of the aerial, the so-called radiation resistance. The wave resistance ζ of a Lecher system may very well be considered as a one-dimensional radiation resistance.

If we have a Lecher system of finite length and connect the two wires at the right end by a pure resistance ζ, the Lecher system behaves as if it extended to infinity on the right. The relation between V and i at the position of the terminating resistance is exactly as in a wave moving to the right on an infinitely long Lecher system. If, however, the terminating resistances have a value differing from ζ there is a different relation between V and i. This deviating relation is not realized by a wave moving toward the right, but only by the superposition of a wave moving toward the right and a wave moving toward the left, which means that reflection takes place.

Velocity of propagation and wave resistance

For several cases of practical importance we shall specify more exactly the self-induction L^I and the capacity C^I per unit of length, in order to calculate from them the velocity of propagation and wave resistance of this Lecher system. In the first place we consider two metal strips with a width b and a distance apart d, lying parallel to each other (*fig. 4*). If d is made small compared with b the

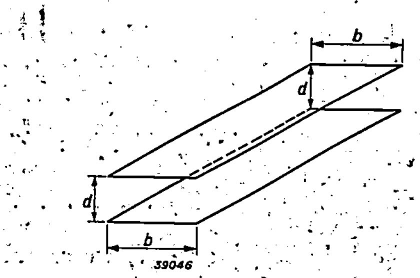


Fig. 4. Lecher system of wide strips. The distance between the strips is d and their width b.

capacity per unit of length in the z direction becomes

$$C^I = \frac{\varepsilon b}{4\pi d}\ \text{cm/cm} = \frac{\varepsilon b}{4\pi d}\,\frac{1}{9\cdot 10^{11}}\ \text{farads/cm}, \quad (12)$$

where ε represents the dielectric constant.

The self-induction L^I per unit of length for this simple case is also easy to find with the usual simplifications. If we assume the magnetic field H between the two strips to be homogeneous, and outside of them equal to zero, then

$$Hb = 0{\cdot}4\pi i \ \text{gauss cm.} \quad (13)$$

If μ is the permeability, the number of magnetic lines of induction which run between the two wide strips per cm in the z direction is $Bd = \mu Hd \times 10^{-8}$, and this by definition is the current i times the self induction L^I per unit of length:

$$iL^I = \mu Hd\cdot 10^{-8}\ \text{amperes henrys} \quad (14)$$

From (13) and (14) it therefore follows that the self induction in henrys/cm becomes

$$L^I = \frac{4\,\pi\mu d}{b}\,10^{-9}\ \text{henrys/cm} \quad (15)$$

The velocity of propagation v for electromagnetic equilibrium disturbances in a Lecher system is according to (12) and (15)

$$v = 1/\sqrt{L^I C^I} = \sqrt{\frac{9\cdot 10^{20}}{\varepsilon\mu}} = \frac{3\cdot 10^{10}}{\sqrt{\varepsilon\mu}}\ \text{cm/sec.}$$

For a medium with a dielectric constant $\varepsilon = 1$ and 1 and a permeability $\mu = 1$ this becomes equal to the velocity of light of 3×10^{10} cm/sec which is universally valid for parallel conductors of any given cross section.

According to formulae (11), (12) and (15) the wave resistance becomes

$$\zeta = \sqrt{\frac{4\,\pi\mu d}{b}\cdot\frac{4\pi d}{\varepsilon b}\cdot 9\cdot 10^2} = 120\pi\,\frac{d}{b}\sqrt{\frac{\mu}{\varepsilon}}\ \text{ohms.}$$

If in an analogous way the capacity and self-induction per unit length are calculated for two concentric conductors with radii R_1 and R_2 (*fig. 5*) one finds

$$C^I = \frac{\varepsilon}{2\ln R_2/R_1}\cdot\frac{1}{9\cdot 10^{11}}\ \text{farads/cm,}$$

$$L^I = 2\mu\ln R_2/R_1\cdot 10^{-9}\ \text{henrys/cm.}$$

The velocity of propagation $v = 1/\sqrt{L^I C^I}$ now becomes $3\times 10^{10}/\sqrt{\varepsilon\mu}$ cm/sec here also, while for the wave resistance the following is obtained:

$$\zeta = \sqrt{2\mu\ln R_2/R_1\cdot 2/\varepsilon\ln R_2/R_1\cdot 9\cdot 10^2} = 60\ln\frac{R_2}{R_1}\sqrt{\frac{\mu}{\varepsilon}}\ \text{ohms.}$$

If we are concerned with two parallel round wires with a radius of R_0 which is small compared with their distance apart

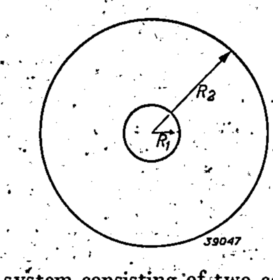


Fig. 5. Lecher system consisting of two concentric cylinders with radii R_1 and R_2.

d (*fig.* 6), the capacity and self-induction per unit of length become

$$C^I = \frac{\varepsilon}{4 \ln \frac{d}{R_0}} \cdot \frac{1}{9 \cdot 10^{11}} \text{ farads/cm,}$$

$$L^I = 4\mu \ln \frac{d}{R_0} \cdot 10^{-9} \text{ henrys/cm.}$$

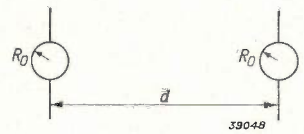


Fig. 6. Lecher system of round wires with a radius R_0 and a distance d between them.

The velocity of propagation is of course again the velocity of light divided by $\sqrt{\varepsilon\mu}$, and for the wave resistance one now finds:

$$\zeta = \sqrt{4\mu \ln \frac{d}{R_0} \cdot \frac{4}{\varepsilon} \ln \frac{d}{R_0} \cdot 9 \cdot 10^2} = 120 \ln \frac{d}{R_0} \sqrt{\frac{\mu}{\varepsilon}} \text{ ohms.}$$

In all these cases it is tacitly assumed that no energy losses occur in the self-inductions and capacities, so that the waves are propagated unweakened along the Lecher system [4]).

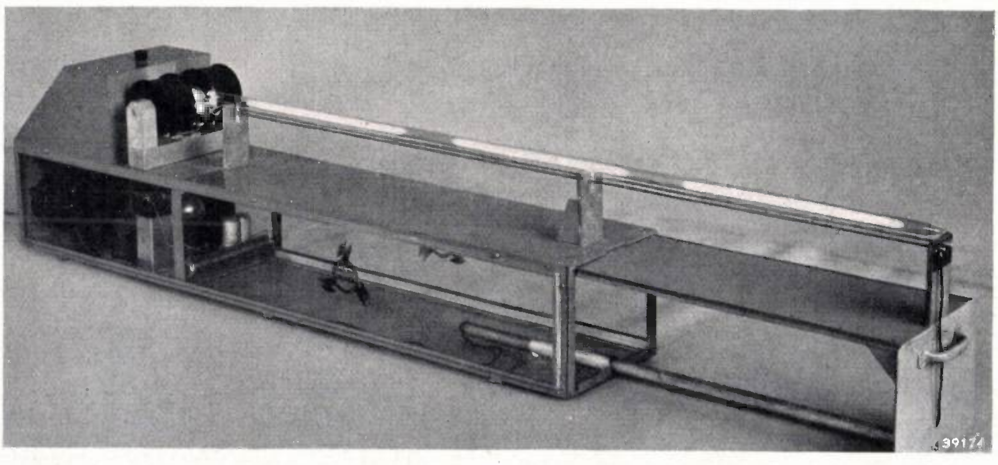
[4]) In the propagation of light and of radio waves in a vacuum no energy losses occur either, but nevertheless the amplitudes decrease in inverse proportion to the distance to the source of radiation, due to the spreading of the energy over larger and larger areas. To this corresponds the fact that upon propagation through space, instead of an expression of the form $f(t-z/v)$, an expression of the form $\frac{f(t-R/v)}{R}$ occurs. The parallel conductors thus keep the radiation energy better together in propagation along a Lecher system.

If the heat losses which occur in a Lecher system are taken into account, the amplitudes become smaller during the propagation along the system. Another cause for such a decrease lies in the fact that due to the finite distance between the two wires some radiation will always occur, since to a certain degree they will act as a loop aerial. We shall return later to such phenomena; for the present we shall continue to describe the behaviour of Lecher systems in so far as the resistance and the radiation loss play only a negligible part.

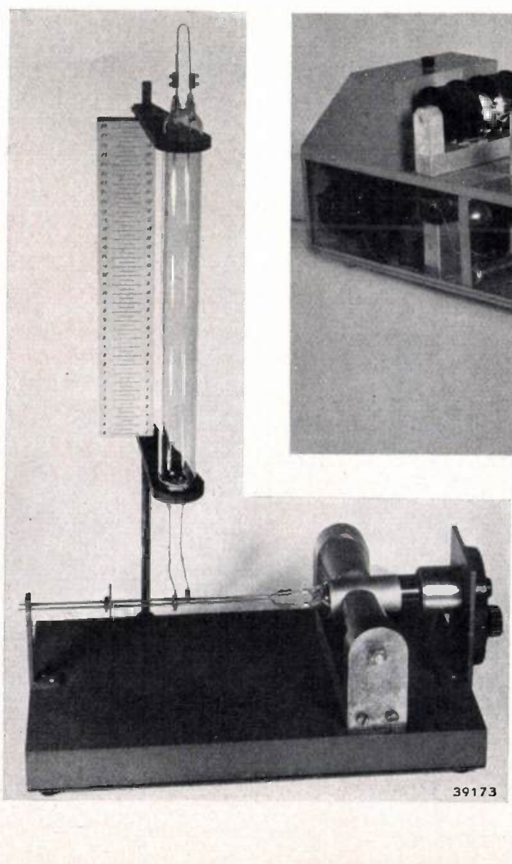
Stationary waves

If a Lecher system is not terminated with its own wave resistance ζ, but with an arbitrary impedance, a wave is entirely or partially reflected as already mentioned and stationary waves occur (see *fig.* 7). Before formulating this process mathematically we shall give a qualitative discussion of the phenomenon for the simple cases of a short-circuited or an open end.

In the left-hand half of *fig. 8* there is a positive and a negative charge which move on the upper and lower wires, respectively, of the Lecher system toward the short-circuited end. This end has the resistance zero, so that the charges e there pass unhindered, and will move back again with the same velocity over the lower and upper wires, respectively, of the system, as is drawn below. The



b



a

Fig. 7. Two demonstration experiments of Lecher systems with stationary waves which are excited by a magnetron oscillator.

In *a* the Lecher system consists of two vertical tungsten wires in a vacuum. At spots where the current density is high (current antinodes) the wires glow. The wave length is about 30 cm.

In *b* a glass tube filled with neon lies above the Lecher system. At spots where the field strength is high (voltage antinodes) the tube is illuminated (wave length about 100 cm).

current i and voltage V corresponding to these moving charges e are also indicated, and it is clear that at this short-circuited end the current is reflected unaltered in the same wire, while the voltage between the two wires is reflected with reversed sign. This corresponds to the fact that at the short-circuited end current can flow, but that there can be no voltage here, since the incident and reflected voltage waves always compensate each other here.

With an open end, to which the right-hand half of fig. 8 refers, an incident charge e is reflected on the same wire, because it has nowhere else to go. The reflected current i is thus reversed in sign, while the voltage V is reflected with the same sign. At an open end, therefore, there may indeed be a voltage, but no current can flow, since the incident and reflected current waves always compensate each other there.

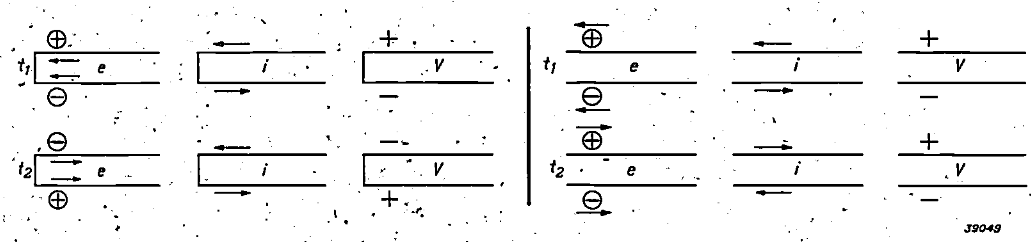


Fig. 8. Diagrammatic representation of the motion of charge e, currents i and voltage V in the neighbourhood of a short-circuited or open end, respectively, of a Lecher system.

If we now consider a Lecher system extending to an infinite length toward the left with a short-circuited end at $x = 0$ (*fig. 9*), on which as a special case a voltage wave $V_1 = \sin\omega\,(t-z/v)$ moves from left to right, a second voltage wave $V_2 = -\sin\omega\,(t+z/v)$ is there reflected toward the left, so that there is a resultant stationary wave V:

$$V = V_1 + V_2 = -2\cos\omega t \sin\omega z/v \quad . \quad . \quad (16)$$

At all points with $\omega z/2\pi v = z/\lambda = n/2$, where n represents a whole number, the voltage is always zero. These points are therefore voltage nodes lying at intervals of a half wave length λ. To the travelling voltage waves V_1 and V_2 according to (9) and (10) belong the travelling current waves

$$\left.\begin{aligned} i_1 &= \frac{1}{\zeta}\sin\omega\left(t-\frac{z}{v}\right), \\ i_2 &= \frac{1}{\zeta}\sin\omega\left(t+\frac{z}{v}\right). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (17)$$

These always reinforce each other at the short-circuited end where a current antinode will occur, which may also be seen directly from the formula for the stationary current wave which occurs as a result:

$$i = \frac{2}{\zeta}\sin\omega t\cos\omega z/v \quad . \quad . \quad . \quad . \quad . \quad (18)$$

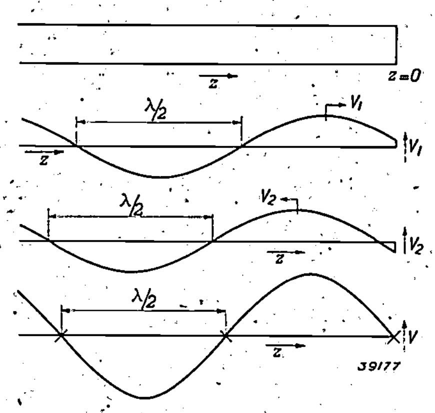


Fig. 9. Reflection of the travelling voltage wave V_1 at the short-circuited end at $z = 0$. The sum of the incident wave V_1 and reflected wave V_2 is the stationary wave V which has nodes at intervals of a whole number of half wave lengths from the short-circuited end of the Lecher system.

If we now consider (16) and (18) somewhat more closely, we see that with the short-circuited Lecher system the current i is 90° behind the voltage V at points close to the short circuit, *i.e.* with sufficiently small negative values of z, while the quotient of the amplitudes of voltage and current amounts to $\zeta\tan\omega z/v$ at any given point. The short-circuit impedance Z_k of a Lecher system with a length l is thus purely imaginary, and is given by

$$Z_k = j\zeta\tan 2\pi l/\lambda \quad . \quad . \quad . \quad . \quad . \quad (19)$$

For a length l which is small compared with the wave length λ, therefore, the short-circuit impedance becomes

$$Z_k = j\zeta 2\pi\frac{l}{\lambda} = j\sqrt{\frac{L^I}{C^I}}\,\omega l\sqrt{L^I C^I} = j\omega L^I l, \quad . \quad . \quad (20)$$

so that a small section of short-circuited Lecher system actually behaves as a self-induction with a value L^I per cm of length. According to formula (19), however, the reactance increases more rapidly than proportional to the length l for the short-circuit Lecher system, so that for a length of $\lambda/4$, for instance, we obtain an infinitely large reactance, which means that at the input of the short-circuited

Lecher system a current node occurs. Such a system can be used as an oscillator curcuit. For a Lecher system with a length between $\lambda/4$ and $\lambda/2$ the reactance is negative, so that it acts as capacity.

In practice use is commonly made of short sections of short-circuited Lecher system as self-induction. We shall give a few such examples here. In the case of radio valves for short waves the capacity is often quite large between points, between which it is desired to introduce an impedance. One then does not make the capacity of the LC circuit to be used variable, because this would mean an extra capacity and therefore a reduction in the impedance attainable. It is then better to tune with a variable self-induction, for which a short section of Lecher system with a movable short-circuiting shunt (*fig. 10*) can very well be used. *Fig. 11* shows the practical application of this device in an oscillator such as is used in the experimental short-wave link [5]) between Eindhoven and Tilburg.

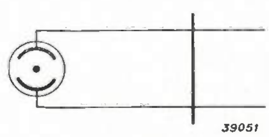


Fig. 10. A short of Lecher system which is provided with a sliding short-circuit bridge, functions as a variable self induction with which it is easily possible to tune the relatively large capacities between the electrodes of radio valves.

Another example is the following. When radio valves are used at very high frequencies it often happens that reflections of the electromagnetic waves occur at the leads through glass. The capacity which is responsible for such reflections can be eliminated by connecting in parallel a Lecher system with a suitably set self-induction (*fig. 12*). A practical example of this is also given as applied in a magnetron oscillator (*fig. 13*).

Just as we have seen in the foregoing considerations that a small section of Lecher system with a short-circuited end behaves like a self-induction, it is also easy to understand that a small section of Lecher system with an open end behaves like a capacity. We here obtain as a special case of the travelling current waves which must compensate each other for $z = 0$ and the travelling voltage waves which are equal at that point:

$$\left.\begin{aligned} i_1 &= +\frac{1}{\zeta}\sin\omega(t-z/v) \\ i_2 &= -\frac{1}{\zeta}\sin\omega(t+z/v) \end{aligned}\right\} \quad . \; . \; . \quad (21)$$

[5]) Philips techn. Rev. 2, 173, 1937.

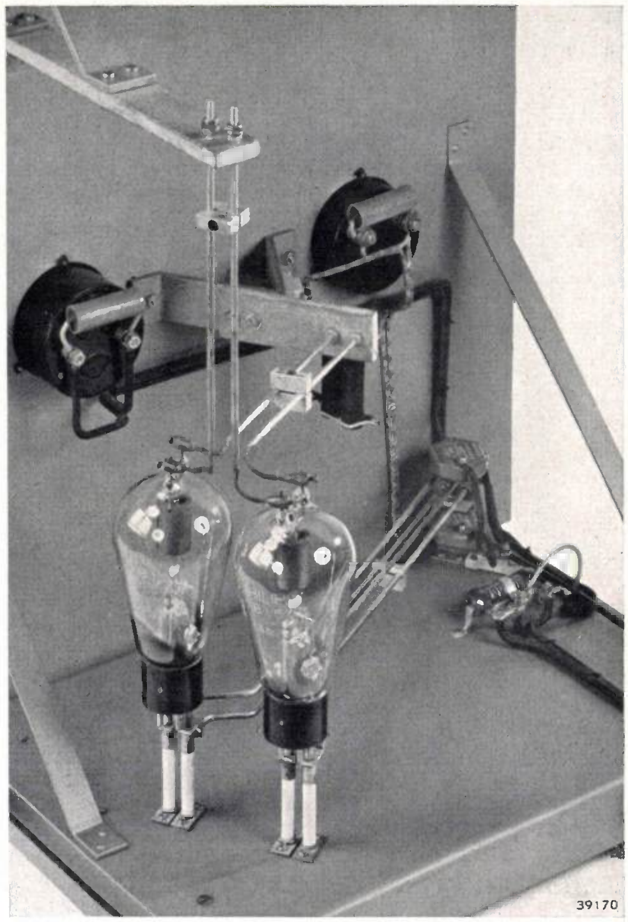


Fig. 11. Triode oscillator in which the admittances of the inter-electrode capacities of anode and control grid are compensated with self-inductions in the form of Lecher systems connected in parallel. With the third Lecher system, which is connected to the cathode, the back coupling is regulated.

and

$$\left.\begin{aligned} V_1 &= \sin\omega(t-z/v), \\ V_2 &= \sin\omega(t+z/v). \end{aligned}\right\} \quad . \; . \; . \; . \quad (22)$$

It then follows from (21) and (22) for the stationary waves that

$$\left.\begin{aligned} i &= -\frac{2}{\zeta}\cos\omega t\,\sin\omega z/v, \\ V &= 2\sin\omega t\,\cos\omega z/v. \end{aligned}\right\} \quad . \; . \quad (23)$$

If we compare (23) with (16) and (18) we see that with the open Lecher system the stationary waves

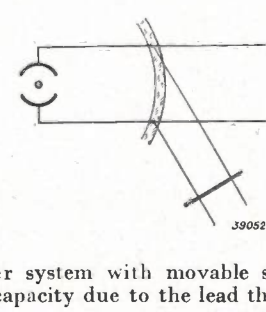


Fig. 12. Lecher system with movable short-circuit bridge for tuning the capacity due to the lead through the glass.

are shifted just a quarter of a wave length with respect to those with the short-circuited Lecher system.

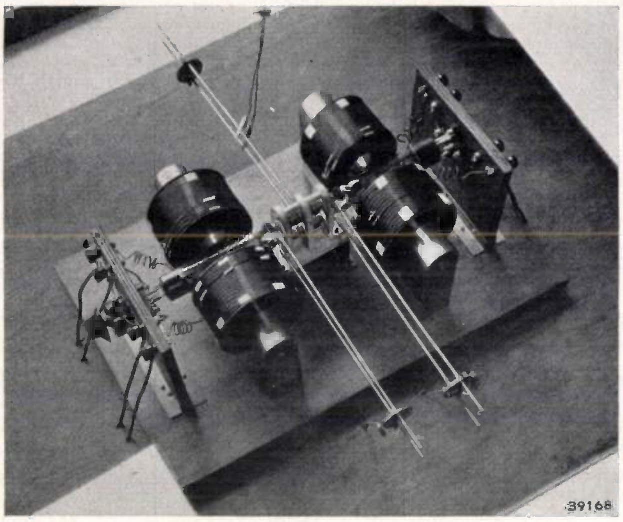


Fig. 13. Magnetron oscillator with two leads through glass, tuned by Lecher systems. The third Lecher system forms the anode impedance of the two valves.

Furthermore, it follows from (23) that with the open Lecher system at points with small negative values of z the current leads the voltage by exactly 90°, so that the open impedance Z_0 is negative imaginary and becomes equal to

$$Z_0 = -j\zeta \text{ c tan } 2\pi \frac{|z|}{\lambda}, \quad \ldots \quad (24)$$

cot

where $|z|$ represents the absolute value of z, *i.e.* the length l of the Lecher system. Z_0 becomes equal to zero at an uneven number of quarter wave lengths from the open end, which means that there the voltage nodes are situated. At an even number of quarter wave lengths from the open end, on the other hand, Z_0 is infinitely large, which corresponds to the current nodes. For a short section of open Lecher system the impedance Z_0 is now better written as

$$Z_0 = -\frac{j\zeta}{\tan 2\pi l/\lambda} \approx -j\sqrt{\frac{L^I}{C^I}}\frac{1}{\omega l\sqrt{L^I C^I}}, \cdot (25)$$

so that we immediately see its behaviour as capacity C^I per unit of length:

$$Z_0 = \frac{1}{j\omega C^I l} \quad \ldots \ldots \quad (26)$$

If we multiply the short-circuit impedance Z_k and the open impedance Z_0 by each other, formulae (19) and (24), we obtain the following remarkable relation:

$$Z_k Z_0 = \zeta^2, \quad \ldots \ldots \quad (27)$$

which is valid for every length of the Lecher system. The wave resistance is thus always the geometric average of short-circuit and open impedance, and may therefore be determined by measuring the two as is often done in practice.

The difference in behaviour between a Lecher system terminated by its wave resistance on the one hand, and a short-circuited or open Lecher system on the other has now been thoroughly dealt with. In the first case current and voltage are everywhere in phase and the energy is propagated unchanged along the Lecher system in order finally to be converted into heat in the pure resistance which terminates the system. In the other case current and voltage are everywhere shifted 90° relatively in phase, and the energy thus oscillates continually back and forth between the electrical and the magnetic field which must alternately be built up and broken down by the stationary oscillations of voltage and current. The Lecher systems with travelling waves may be used for the transfer of electromagnetic energy, while those with stationary waves are often used at high frequencies in such places where at lower frequencies an ordinary LC circuit can be used.

Lecher system as coupling line

If we have a Lecher system closed at point *1* by any given impedance Z_1, the (travelling or stationary) electromagnetic waves at that extremity satisfy the following condition:

$$\frac{V_1}{i_1} = Z_1.$$

At a point *2* at a distance of half a wave length from *1* (*fig. 14*) all components of current and voltage are equal and opposite to those at *1* at every moment, and this means that the impedance Z_2, which can be measured at point *2* of a Lecher system which is closed by Z_1 at *1*, is simply the same as this terminating impedance Z_1.

$$Z_2 = \frac{V_2}{i_2} = \frac{-V_1}{-i_1} = Z_1.$$

In the same way of course for a Lecher system of a whole number of half wave lengths which is

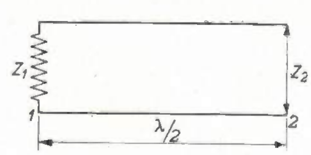


Fig. 14. A Lecher system closed at one end *1* by any given impedance Z_1 exhibits at point *2*, which lies a half wave length from *1*, an impedance $Z_2 = Z_1$.

terminated by any given impedance Z, the impedance at its input is always equal to this Z. From this we see, therefore, that a Lecher system ~~of a~~ of a whole number of half wave lengths is suitable for including in a given circuit the two terminals of an impedance which would otherwise be inaccessible, without the value of the impedance being apparently changed in this circuit.

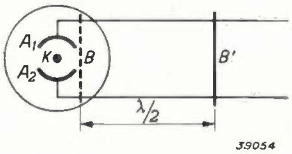


Fig. 15. If the tuning bridge B cannot be put into position because it would fall inside the valve, the same result can be obtained by means of a bridge B' a half wave length farther away.

When radio valves are used on short waves it often occurs that in order to obtain resonance in the anode circuit the bridge B should actually be introduced inside the valve. Instead of this, however, there is no objection to introducing the bridge B' half a wave length away on the supply lines, as is represented diagrammatically in *fig. 15*. This case is illustrated in fig. 7*b*. In order to make more than one half wave length visible the bridge is introduced two half wave lengths farther away than in fig. 15. *Fig. 16* represents such a connection between the control grids of a double pentode.

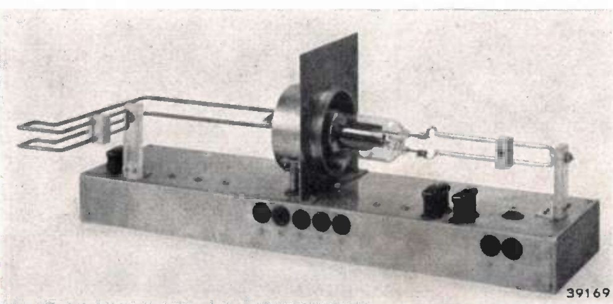


Fig. 16. Double beam tetrode of the wireless telephone connection Eindhoven-Tilburg. Between the two control grids is a tuning bridge such as is shown diagrammatically in fig. 15.

Lecher system as impedance transformer

We shall now consider Lecher systems which are a quarter wave length long and terminated at the end *1* by any given impedance Z_1. We are interested in the impedance Z_2 at the other end *2* which is separated from the closed end *1* by a distance $\lambda/4$. It is found that the product of these two impedances is equal to the square of the wave resistance:

$$Z_1 Z_2 = \zeta^2 \quad . \; . \; . \; . \; . \; . \; . \; . \quad (28)$$

If an impedance Z_2 is included in any given circuit by means of a Lecher system of an uneven number of quarter wave lengths, it behaves as a directly connected impedance with the value $Z_1 = \zeta^2/Z_2$. Thus in general a Lecher system of the length $(2n + 1)\ \lambda/4$ acts as an impedance transformer, and it may be used to connect two circuits of which it is desired that the loading impedance of the first shall differ from the input impedance of the second.

Such a case occurs for example when an aerial must be adapted to a supply line. If a given supply line is present, the question may be asked as to the resistance with which this must be terminated so that the losses will be as small as possible. The line should then be terminated by its own wave resistance ζ so that travelling waves will occur. The aerial to which the energy must be conducted will, however, in general have a different radiation resistance than ζ, so that it is necessary to place an impedance transformer between them. If for example the supply line has a wave resistance of 300 ohms, while the freely radiating dipole aerial of $^1/_2$ wave length has a radiation resistance of 73 ohms, a quarter wave length transformer with a wave resistance of $\sqrt{300 \times 73} \approx$ ohms should be placed between them.

The adaptation between supply line and resistance to be supplied R (aerial for instance) might also be achieved without an impedance transformer, by so changing the distance between the wires of the supply line that ζ becomes equal to R. The losses would, however, increase, since with a constant current amplitude at the end of the supply line the stationary waves pass over into travelling waves, and the average current amplitude along the supply line becomes greater. One may not therefore say that the absence of stationary waves on supply lines is desirable in all circumstances.

Resonance resistance and quality factor

We shall now take the energy losses into account in such a way that we leave the current distribution as if there were no losses.

For a short- circuited Lecher system of a quarter wave length which is used as a circuit element, we wish to know how its impedance changes with the frequency. In the article already referred to [6]) we have made use of the quality factor Q to characterize the behaviour of resonance circuits for high frequencies. This is defined as the quotient of the resonance frequency ω_0 and the resonance width $\Delta\omega$, and may in general also be considered as 2π times the quotient of field energy and heat developed

[6]) Philips techn. Rev. 6, 217, 1941.

per period:

$$Q = \frac{\omega_0}{\Delta\omega} = 2\pi \frac{\text{field energy}}{\text{heat developed per period}} \cdot \quad (29)$$

From this it now follows for the short-circuited Lecher system of a quarter wave length that

$$Q = \frac{\omega L^I}{r^I}, \quad \ldots \ldots \quad (30)$$

when r^I represents the ohmic resistance per cm length.

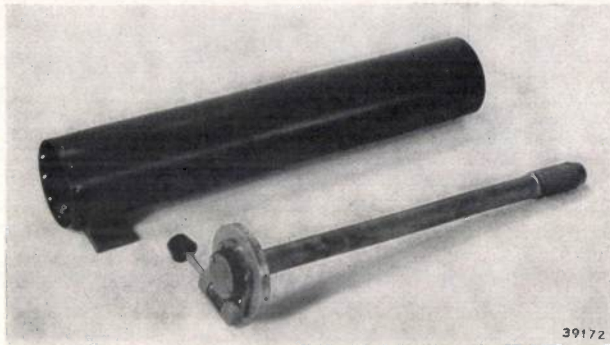
The resonance resistance R is by definition equal to the quotient of the mean square of the voltage V and the heat W developed per second. In the case of a Lecher system of a quarter wave length one calculates for this

$$R = \frac{L^I}{C' r^I \lambda/8} = \frac{4}{\pi} Q \zeta. \quad \ldots \quad (31)$$

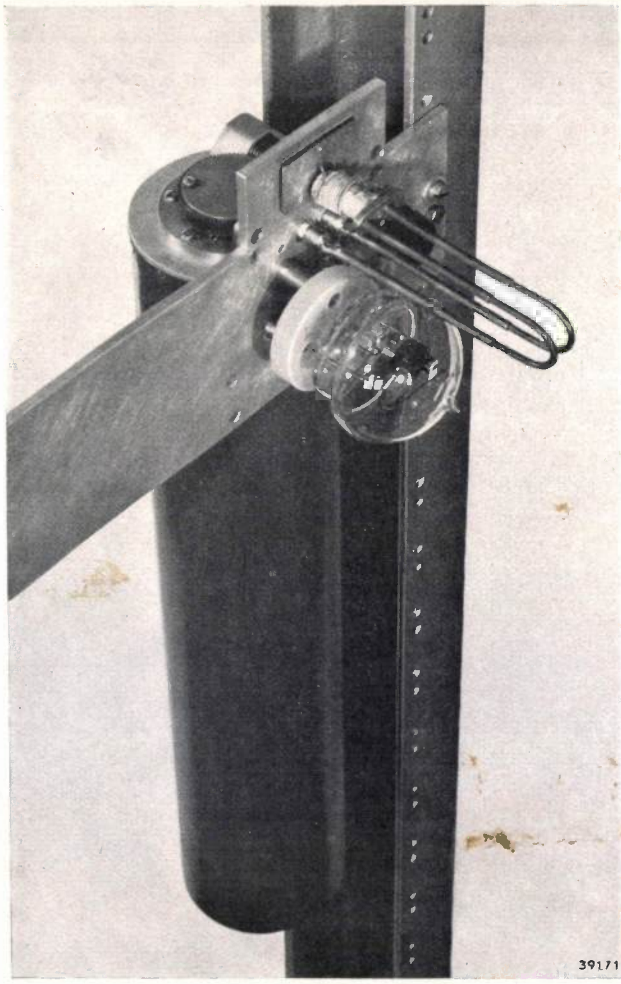
If it is desired to use a Lecher system of a quarter wave length in an amplifier circuit, the resonance resistance must be large, and at a given sharpness of resonance, *i.e.* a given ε, care must therefore be taken that the wave resistance ζ is large. In the three cases of Lecher systems already mentioned (figs. 4, 5 and 6) one finds for the resistance r^I per cm, the quality factor Q and the resonance resistance R, the formulae given in the table below, in which δ stands for the depth of pentetration of the skin effect.

	r^I	Q	R
1)	$\frac{240\pi^2\delta}{b\lambda}$	$\frac{d}{\delta}$	$\frac{480\,d^2}{b\delta}$
2)	$\frac{60\pi\delta}{\lambda}\left(\frac{1}{R_1}+\frac{1}{R_2}\right)$	$\frac{2}{\delta}\frac{R_1 R_2}{R_1+R_2}\ln\frac{R_2}{R_1}$	$\frac{480}{\pi\delta}\frac{R_1 R_2}{R_1+R_2}\left(\ln\frac{R_2}{R_1}\right)^2$
3)	$\frac{120\,\pi\delta}{R_0\lambda}$	$\frac{2R_0}{\delta}\ln\frac{d}{R_0}$	$\frac{960}{\pi}\frac{R_0}{\delta}\left(\ln\frac{d}{R_0}\right)^2$

In general the quality factor Q is proportional to the quotient of volume and surface of the oscillation circuits. In the last case, two round parallel wires, a volume of the system cannot well be given, but this is possible in the first and second cases. Particularly in the case of the Lecher system of two parallel strips it is immediately evident that from $Q = d/\delta$ it follows that Q is equal to $2/\delta$ times the quotient of volume and surface of the oscillation circuit, which is also confirmed for two concentric conductors whose radii R_1 and R_2 do not differ too much. For two concentric cylinders whose radii are very different the current density in the inner cylinder is much greater than in the outer, and consequently this general formula no longer holds exactly. From the second line of the table it is possible to calculate those ratios of the radii of the two concentric cylinders for which the quality



a



b

Fig. 17. Concentric open Lecher system of slightly more than a quarter wave length, which serves as a practically loss-free small self-induction for the stabilization of the high-frequency oscillations excited by a radio valve. The inner rod is made of quartz to prevent thermal variations in length, and is surrounded by a covering of copper in which there is an accordionlike connecting piece. In a the components may be seen, in b the application of the system in a short-wave transmitter of 1.2 m wave length.

factor Q and the resonance resistance R are as large as possible. One finds then $R_2/R_1 = 3.7$ and approximately 9, respectively.

Lecher system as a stabilizer

The capacities between the electrodes of transmitting valves do not remain entirely constant while the valve is in use, and since these capacities form part of the tuned circuit the frequency of the oscillation excited varies. This is avoided by the use of connections in which a large capacity is connected in parallel with such variable capacities so that the influence of the latter becomes much smaller. The oscillation circuit is then further tuned to resonance by a self-induction which is in parallel with these large capacities and which must therefore be small. In addition to cavity resonators, concentric open Lecher systems can also very well be used for this purpose. Such Lecher systems have a very high quality factor, and with a length of slightly more than a quarter wave length, they form a small, practically loss-free self-induction as may be seen directly from (24). Such a concentric Lecher system to be used as stabilizer in short-wave transmitters is shown in *fig. 17*.

Radiation losses

In the foregoing discussions we have only taken into account the ohmic losses in the Lecher system by ascribing to it a resistance r^I. Radiation losses may, however, also occur, and we shall discuss them briefly in conclusion. In practice radiation of energy is often avoided by shielding the Lecher system by means of a metal tube, If, however, this has not been done, then with a short-circuited Lecher system of a quarter wave length with a distance d between the conductors, the energy radiated per second is approximately

$$P = 1200\, i_0^2 \frac{d^2}{\lambda^2} \text{ watts}, \quad (32)$$

where i_0 is the maximum current in the bridge. The same amount of heat would be developed in an equivalent resistance r_{eq} per cm distributed uniformly along the Lecher system:

$$i_0^2\, r_{eq}\, \lambda/8 = 1\,200\, i_0^2 \frac{d^2}{\lambda^2},$$

so that

$$r_{eq} = \frac{9\,600\, d^2}{\lambda^3}. \quad (33)$$

For a Lecher system of a quarter wave length which consists of round wires with a diameter of 2 mm stretched at a distance of 1 cm away from each other, the following equivalent radiation resistance is found at a wave length of for instance 75 cm:

$$r_{eq} = 0.025 \text{ ohms per cm}, \quad (34)$$

which is of the same order of magnitude as the ohmic resistance r^I per cm for this system.

MERCURY LAMPS FOR USE IN MAKING HELIOGRAPHIC PRINTS

by A. A. PADMOS and J. VOOGD.

621.372.3 : 778.16

The requirements are studied which must be made of lamps for use in making heliographic prints. A mercury discharge lamp with atmospheric pressure constructed for this purpose is briefly described. In conclusion the best method of using this lamp in heliographic printing machines is indicated.

Introduction

With the help of the heliographic printing process it is possible to make reproductions quickly in natural size of tracings, pages of typing and the like. The method of the "contact print" is usually applied. The light passes through the "original" which is to be reproduced and then falls upon the printing paper which is placed behind it. In simple printing frames such contact prints are made one at a time as in making copies in ordinary photography. In larger printing machines the printing paper, together with the original, is carried by a transport band along the back of a bent glass surface. The operation of such a machine is represented diagramatically in *fig. 1.* A source of light L illu-

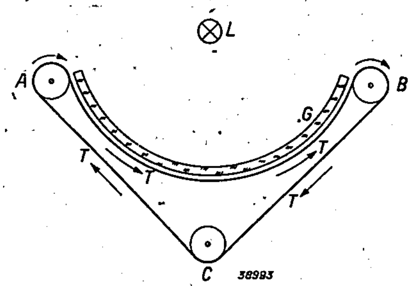


Fig. 1. Principle of the method of operation of a heliographic printing machine. L source of light, G glass plate, T transport band running over the cylindrical rollers A, B and C.

minates the cylindrically bent glass plate P. The transport band T runs along the back of this plate. At the roller A the original together with the printing paper is inserted between the glass and the transport band. The speed with which the transport band T carries the paper along the glass then determines the time of exposure. At the roller B the original and the print are taken off the machine and the print is then developed. With the help of the roller C the transport band may be put under more or less tension.

The success of this method is chiefly due to the fact that cheap kinds of printing paper can be obtained, which because of certain properties are especially suitable for use on a technical scale. One of these is the fact that with these kinds of printing paper the photochemical reaction is caused by the action of violet radiation. This radiation occurs to such a small extent only in the ordinary illumination of work rooms that no special precautions need be taken for protecting the paper from the light in order to prevent premature photochemical reactions.

For the rapid production of prints an intense violet illumination must be provided. The heliographic printing process would not have become so common if there had been no sources of radiation available whose properties are well adapted to those of the printing paper. It was only because of this that it was possible to use the short exposure times, without which the modern heliographic printing industry would be unthinkable.

Requirements for the printing lamp

Originally heliographic prints were made exclusively with daylight. The making of a single print then required an exposure time of many minutes. Practically no improvement could be obtained by using electric lamps for printing, since they also provide only very little violet radiation.

The introduction of the carbon arc lamp, however, meant an important advance, since this lamp gives a reasonable intensity of violet radiation so that with it printing speeds are reached which vary from a half to several minutes, depending on the nature of the original which is to be printed and on the kind of printing paper used.

There are, however, disadvantages connected with use of the carbon arc lamp which justify the search for a different type of light source for printing machines. For example, carbon arc lamps require much care since new carbons must be regularly inserted and the lamp glasses must be cleaned. Furthermore they often fail to burn quietly, while due to the lack of uniformity in the light distribution it is sometimes necessary to move the arc. The development of mercury vapour lamps made it possible to avoid all these difficulties. These lamps require no upkeep, burn very regularly and can be so constructed that they give a suitable light distribution.

If we briefly examine the requirements which are made of a printing lamp, the first will be that of a short exposure time. This means that lamps must be used which have a high power and which moreover give a large part of their radiation in the violet. As to the total output of the lamps in a printing machine, it will be limited by the maximum temperature which is permissible on the glass plate and transport band. In large printing machines with carbon arcs the lamps consume 6 to 10 kW. The total power of the mercury vapour lamps which must be installed in such a machine may therefore not be greater than this order of magnitude.

In the second place the efficiency and the life of the mercury vapour lamps should justify their use as printing lamps economically also, while in the third place the dimensions of the lamp should be chosen so that a uniform light distribution is obtained. The linear form of the mercury discharge lamp is particularly suitable for this purpose. For a printing machine with a transport band 1 m wide a satisfactorily uniform light distribution can be obtained with two lamps 50 cm long in a line.

Finally it is necessary that the current or voltage should not be unreasonably high upon ignition or during use. It has indeed been found possible to construct a mercury vapour lamp which satisfies the requirements here mentioned and which has already proved its value as a printing lamp in practical use.

Construction of the printing lamp

In the development of the printing lamp the foundation was a mercury discharge lamp with a pressure of about 1 atmosphere. Since during use the temperature of the surroundings and the cooling of the lamp are practically constant a single layer glass wall was sufficient. The distance between the oxy-cathodes is 55 cm, while the external diameter of the lamp is 3 cm. The length of the light column of 55 cm was chosen in order to be able to illuminate a width of 100 to 110 cm uniformly with two lamps placed end to end or overlapping slightly. The lamp consumes a power of 1 900 W at a current of 8.7 A and a working voltage of about 240 V. In series with a suitable choking coil the lamp can be connected to an A.C. voltage of 380 V. With the help of a specially constructed leakage transformer the lamp can also be connected to 220 V. Lamp and series apparatus consume a total of about 2 kW. In *fig. 2* this printing lamp with choking coil is shown.

In order to characterize the spectral energy distribution of the lamp, the energy current density I_λ in erg/cm^2 sec in the perpendicular plane bisecting the axis of the lamp at 100 cm distance from the axis for different wave lengths is given in *table I*. In the last column of table I the contribution of the different wave lengths to the photochemical reaction in the printing paper is given. For this purpose I_λ is multiplied by the spectral sensitivity V_λ of the paper and by the spectral transmission d_λ of the glass plate. The values in the table are given on a relative scale such that for the wave length 4 047 Å the value of 100 is obtained.

The speed at which prints can be made with this mercury lamp is practically the same as with an arc lamp of the same power. The exposure times necessary for making a print vary between about 10 and 15 sec.

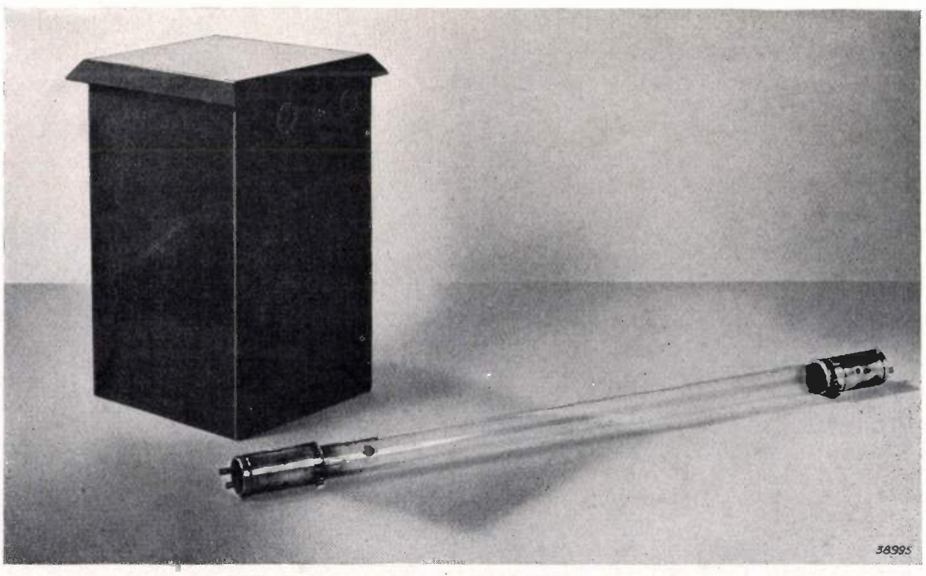


Fig. 2. Printing lamp with an arc length of 55 cm with corresponding choking coil.

Table I

λ in Å	I_λ in erg/cm² sec at a distance of 100 cm	$I_\lambda\ V_\lambda\ d\lambda$ in a relative scale
5 770-5 791	7 920	0
5 461	6 360	0
4 358	4 930	132
4 047	2 520	100
3 655	3 490	75
3 342	70	0.20
3 130	105	0
3 012	20	0

Use of the printing lamp

The mercury lamps here described have of course quite a different light distribution from that of the carbon arc lamps used formerly. In the end the use of the new printing lamps will also lead to a different construction of the printing machine, but it has, nevertheless, been found quite possible to replace the carbon arcs in existing machines by mercury lamps.

Fig. 3 shows such a printing machine which is now equipped with mercury lamps. By placing mirrors at the ends of the lamps perpendicular to the axis of the tube the distribution of the illumination could be made more uniform and, moreover, the necessary exposure time could be shortened. In practice, however, it is found that the refitted printing machine of fig. 3 is quite satisfactory; the advantages of the mercury discharge over the carbon arc are fully exploited, while with the same power of the lamps prints can be made at the same speed.

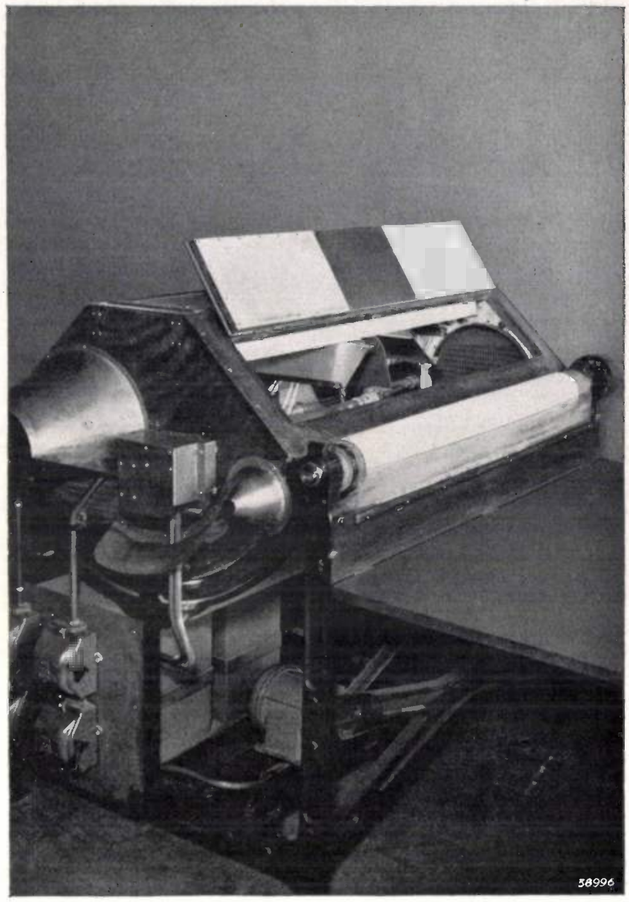


Fig. 3. Printing machine in which the carbon arcs have been replaced by mercury lamps. To the left below the machine may be seen the choking coils belonging to the mercury lamps.

VOL. 6. No. 9, p. 253-284 SEPTEMBER 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

SEVERAL TECHNICAL PROBLEMS IN THE DEVELOPMENT OF A NEW SERIES OF TRANSMITTER VALVES

by E. G. DORGELO.

621.396.615

The use of shorter and shorter radio waves involves a steady decrease in the dimensions of the transmitter valves. When it is at the same time desirable not to decrease the energy dissipation in the valve, various parts of the valve become relatively more heavily loaded. An indication is given in this article of how several technical problems connected with this are solved in the case of a new series of Philips transmitter valves.

Introduction

Under the influence of the rising standard of performance required in the field of television, a great demand has arisen in recent years for transmitter valves which can work with high efficiency on the wave lengths of 5 to 6 m used in television. The existing types could to some extent be used for such wave lengths, but usually worked at low efficiency, and, moreover, they often involved great difficulties in the design and adjustment of the transmitter.

In connection with this an entirely new series of transmitter valves has been developed, which is suitable not only for long waves but also for the short waves mentioned above. In appearance this type of valve differs from the older types mainly in its much smaller size. This small size was necessary on the one hand to diminish the transit time effects, and on the other hand to make possible a better adaptation to the transmitter by smaller electrode capacities and smaller self- and mutual inductions in the connections.

The series of valves now developed consists of four triodes of about 250, 600, 1 200 and 2 500 W telegraphy output [1]) respectively, three pentodes of about 200, 500 and 1 000 W and a push-pull pentode also of about 1 000 W (see *fig. 1*).

The application of the push-pull principle was justified by the consideration that otherwise at the wave lengths of a few metres many of the advantages of the pentode connection are lost (for a detailed discussion of the advantages and disadvantages of triodes and pentodes we refer to an article published earlier [2])). While at longer wave lengths screen grid and suppressor grids function as shielding cage due to their constant potential, and thereby make special decoupling measures (neutrodyning) unnecessary, at very short waves this is no longer true. The self-induction of the connections forms such a high impedance at these frequencies that the screen-grid alternating current causes the potential of the screen grid to vary appreciably.

A solution of this is to connect two valves in push-pull connection and supply the corresponding electrodes together through a single line, so that the alternating currents cancel each other in this line. There then occurs no A.C. voltage along these connections. Between the common connection and each of the electrodes, however, there remain connections which are not in common. These connections may also still have too much impedance at very short waves, so that recourse must then be had to reducing their impedance with the help of series resonance. The filament connection also must often be tuned in this way (see *fig. 2*).

These precautions are, however, at least with wave lengths of a few metres, not yet necessary when the two electrode systems are assembled in a single bulb

[1]) By "telegraphy output" we mean here the maximum output in class C adjustment. With this arrangement the control grid voltage is so strongly negative that anode current flows for less than half a period.

[2]) J. P. Heyboer, Philips techn. Rev., 2, 257, 1937.

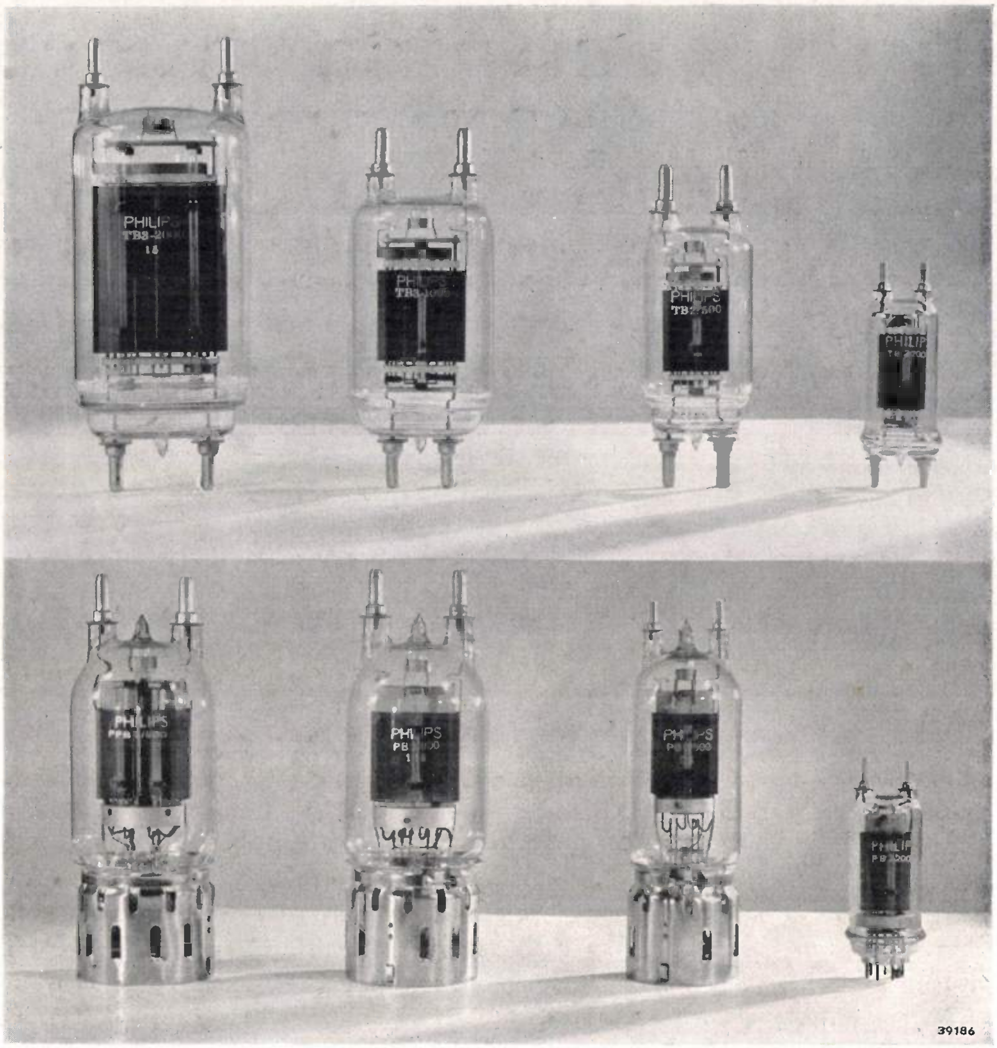


Fig. 1. Photograph of a series of new transmitter triodes and pentodes which were developed on the principles set forth in this article. The output in telegraphy adjustment, class C, for the series of pentodes (lower row) is, from left to right: 1 000, 1 000, 500, 200 W, and for the series of triodes (upper row) it is 2 500, 1 200, 600, 250 W.

and joined with the shortest possible connections. A particularly compact solution is obtained by surrounding the two cathodes and control grids with a single screen grid and suppressor grid (*fig. 3*). This concept is realized in the push-pull pentode PPB 3/800 which can work entirely without extra tuning to the shortest wave lengths which it can reach (2.5 metres).

For the fundamental requirements which every transmitter valve must satisfy we may refer to an earlier article in this periodical [3]). In addition to the points touched upon here, which are connected with form and manner of construction of the modern transmitter valves, the requirement of small size also raises technical problems which will be discussed in the following sections.

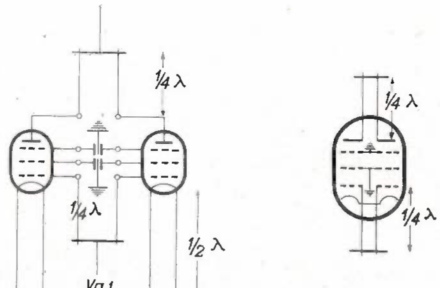


Fig. 2. *a*) Diagram of two pentodes in push-pull connection. The parts of suppressor and screen grid connection not possessed in common are tuned with the help of variable series capacities. The regulation of the filament impedance is by means of Lecher systems of the length $^1/_2\ \lambda$. As oscillation circuits in anode and control grid circuits Lecher systems are also used (length $^1/_4\ \lambda$).

b) In the push-pull pentode PPB 3/800 only the grid and anode circuit are tuned. The filaments are connected with very short connections inside the valve. Furthermore there is only one common suppressor and screen grid.

---

3) H. G. Boumeester, Philips techn. Rev., 2, 115, 1937.

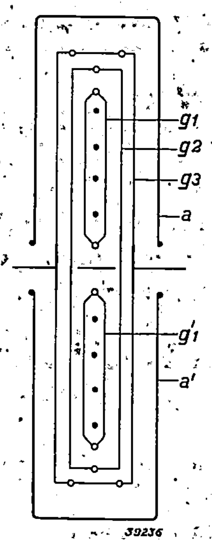


Fig. 3. Cross section of the electrode system of the push-pull pentode PPB 3/800. The shielding between the two halves of the anode prevents electrons from the right-hand cathode from reaching the left-hand anode, and *vice versa*.

Fundamental differences between the modern and the older types of valves

We shall here first mention those parts of a valve to which special attention must be paid, and later discuss two of them in more detail.

a) Grids and anode

Small dimensions mean small surfaces, and thus with a given power a high specific dissipation. It is therefore necessary to use for the electrode system a material which can withstand high temperatures. Nickel and iron are less suitable because of their fairly great volatility; molybdenum, tungsten and tantalum are better. In the new series of Philips valves molybdenum has been used exclusively for anode and grids, and all connections are clinched with molybdenum rivets or welded directly.

b) Cathode

The high working temperature, which in the case of the anode may amount to 800—900 °C, prevents the use of anything but pure or thoriated tungsten for the cathode. Oxide cathodes would be too much overloaded due to heating from the other electrodes. Considering the high emission of thoriated tungsten compared with pure tungsten (about 70 mA/W *versus* 6 mA/W) the former has been chosen. With a relatively low filament power a good anode current can now be obtained, so that these valves can deliver a satisfactory output with a fairly low anode voltage (3 000 V or lower). Especially on the shortest waves where the circuit losses become very great at high voltage this is important (see *fig. 4*).

c) Bulb

The fact that great power is dealt with in a small electrode system necessitates making the covering of the valve, the bulb, either very large in diameter or of a kind of glass which does bot easily soften. When hard glass is used the bulb can be made quite narrow, so that short leads with slight self- and mutual induction are obtained. The use of this glass at first presented difficulties because of the fact that the existing kinds of hard glass became slightly conducting, especially at the high working temperature here prevailing, so that particularly the fused-in leads were electrolytically attacked. By making use of newly developed glasses and a fusing-in technique borrowed from the quartz lamps, however, it was possible entirely to overcome these difficulties, as will be described in the following.

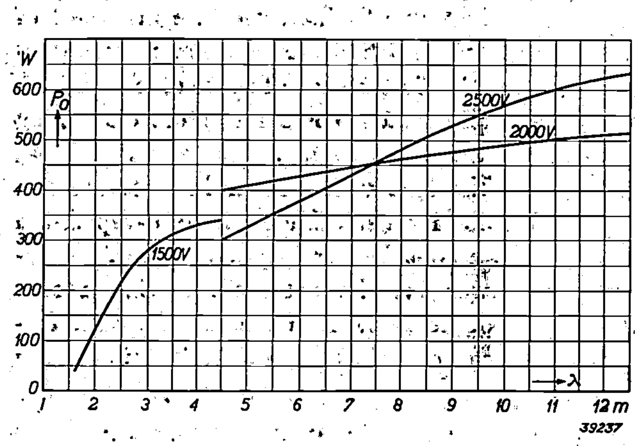


Fig. 4. Output of a PPB 3/800 valve as a function of the wave length. It may be seen that it is more advantageous to lower the anode voltage to a definite limit at short wave lengths, since the losses in the circuits of the transmitter increase too rapidly with the voltage. In connection with this effect it is desirable to construct valves for short wave lengths for a relatively low anode voltage.

d) Getter

Another result of the high temperature was the difficulty of maintaining a sufficiently high vacuum during use. Ordinary getters such as barium and magnesium could not be used here for various reasons. Aside from the great hindrance to the heat radiation formed by the mirror deposited on the bulb wall, the mirror also constitutes an undesired and badly reproducible capacity with respect to the electrodes. At a bulb temperature of 300-350 °C the vapour pressure of barium and magnesium is also already so high that one may scarcely speak of a "vacuum".

Zirconium, whose getter properties have already been discussed in this periodical [4]), lacks the dis-

[4]) J. D. Fast, Philips techn. Rev. 5, 217, 1940.

advantages mentioned. It can easily be deposited on the electrodes in powder form and thereby automatically takes on the temperature necessary for satisfactory functioning.

In the following we shall go into somewhat more detail about the bulb and the getter.

The bulb and the electrode leads through the glass

As already mentioned, electrical conduction easily occurs in the glass of the bulb, especially between the fused-in leads. Conduction in glass, like the electrolysis of liquids, is based upon the movement of ions. In the conduction in glass substances are also deposited at the poles. The presence of these substances results in general in a lessened reliability of the fused-in leads. The oxide layer usually present on the metal is discoloured, gas bubbles are formed in the boundary layer and after some time the valve begins to leak. As a result of di-electric losses which always accompany the electrical conduction in the case of bulbs of transmitter valves, this phenomenon becomes more disturbing the shorter the wave length.

Since all the ions which occur in glass are not equally mobile, by a suitable choice of composition of the glass a considerable improvement of the insulating power can be attained. During recent years types of glass have been successfully manufactured whose electrical conductivity is many thousand times smaller than that of ordinary hard glasses (*fig. 5*).

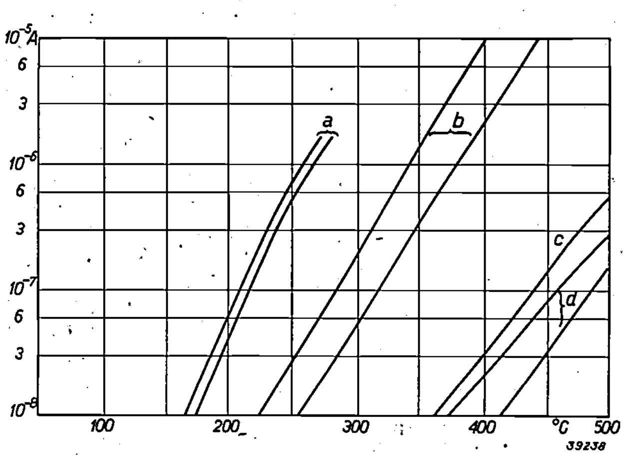


Fig. 5. Electrolytic conductivity of glass as a function of the temperature. In a piece of glass two wires of 1 mm diameter are fused in over a length of 1 cm at a distance apart of 1 cm. Between the wires a D.C. voltage of 100 V is applied. The current increases approximately according to a power of e with the temperature. Four groups of glass can be distinguished:
a) Borosilicate glass. Softening point 525-550 °C. Resistance low.
b) Soft glass (lead glass). Softening point about 450 °C. Resistance fairly high.
c) Hard glass. Softening point 700-1000 °C. Resistance high.
d) "Electrolysis-free" glasses. Softening point 550-600 °C. Resistance very high.

Another great improvement was the introduction of bare metal leads through glass. By means of the correct technique of fusing in, in combination with the use of suitable kinds of glass, it was found possible to fuse in leads of molybdenum and tungsten entirely free of oxide. Such bare leads are distinguished from the earlier leads by the fact that even at working temperatures of 400 °C and more they remain absolutely vacuum tight.

Leads through glass of this type are used in the pentodes of the series here described for the base leads (filament, control, screen and suppressor grids).

The top leads (anode and a second connection of the suppressor grid) have a considerably larger high-frequency current to withstand, since the output circuit is connected to them. Of themselves the bare molybdenum or tungsten leads through glass would be quite satisfactory here; it is, however, difficult to fasten a suitable binding post arrangement to them. Direct connection is out of the question because of the brittleness of fused-in molybdenum or tungsten.

A much stronger arrangement is obtained by the use of heavy copper pins which are fused to the bulb by means of a "fernico" ring soldered to them. These leads can easily withstand 50 A or more.

In the triodes of this series copper with fernico is also used for the base connections (filament).

The zirconium getter

Pure zirconium has a very strong tendency, depending on the temperature, to bind gases. When used in powder form a large active surface is obtained while gas is then taken up even at low temperatures. In general it may be said that for binding oxygen and nitrogen red heat is necessary, while hydrogen is taken up only by the non-glowing parts (see the article referred to in footnote [4]).

In a transmitter valve all these temperatures are as a rule present at the same time. Thus in the case of the largest of the valves here described the maximum anode temperature is about 800 °C, the lowest about 300 °C, while connection strips and the like may be still colder. Since a good vacuum is necessary here in the first place for maintaining the cathode emission (tungsten-thorium cathodes are very sensitive to oxygen), it is obvious that the zirconium powder should be placed in the immediate vicinity of the filament.

In the case of the pentodes of this series the screen grid as well as the anode is covered with zirconium. In the case of the triodes also it was natural to make use of such a shielding cage which in this case could only be formed by the control grid. While indeed the control grid of a triode usually has a lower working temperature than the screen grid of a pentode, the getter action is in this case increased by the fact that the zirconium on the control grid has a low electric potential with respect to the surroundings, so that any positive ions present are drawn to it. It was actually found possible to maintain a good vacuum in a triode also by this means.

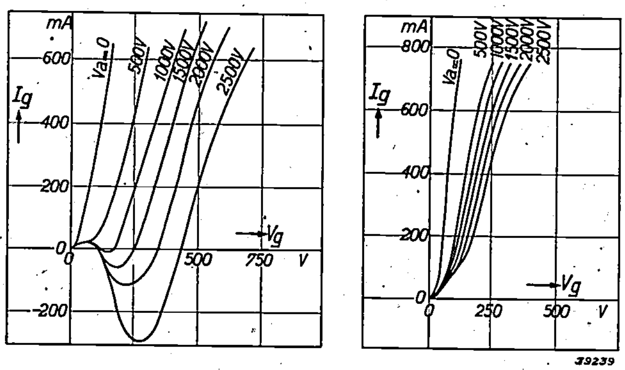


Fig. 6. Grid current characteristics of the valve TB 3/1 000, *a*) without zirconium, *b*) with a very thin layer of zirconium powder on the grid. Because of the fact that the coefficient of secondary emission of molybdenum is greater than unity for electron velocities of several hundred volts, a negative grid current may occur in *a*). In practice, however, a high positive grid voltage is accompanied by a low anode voltage, so that with a well constructed valve the momentary value of the grid current always remains positive.

The covering of electrodes with zirconium has still other results. As was described in the article referred to in footnote [3]), zirconium decreases the secondary emission of the surface upon which it is deposited. Since the total current to an electrode is the sum of the primary emission falling on the electrode and the secondary emission counter to it, influence can be exerted on the total current by covering the electrode with more or less zirconium.

In the case of the control grid the desire is as a rule to make the total current as small as possible. Every increase in the grid current is accompanied by an increase in the excitation power and thus a decrease in the energy amplification. It is therefore best in this case when the secondary emission is as nearly as possible equal to the primary electron current received.

Now in the case of modern transmitter valves this is usually approximately the case even without the deposit of zirconium. They are constructed for low voltage and great current density so that there is so much space charge in the valve that secondary electrons pass through it with difficulty, and the secondary emission current, which of itself is considerably larger than the primary current, is reduced to about the desired value.

Covering the grid with zirconium would in such valves disturb the equilibrium between the currents, and thus have an unfavourable effect. In order to give some idea of the change in grid current caused by covering it with zirconium, several characteristics are shown in *fig. 6*. In the case of the valve with zirconium-covered control grid, used in a high-frequency connection, the grid current was 165 mA and the energy amplification 18 times, compared with 90 mA and 39 times in the case of the ordinary valve. In order to prevent such an increase in the grid current, in the new valves with zirconium getter only those parts of the control grid were covered which are not exposed to a direct bombardment by primary electrons. The parts here involved are the grid rod and in some cases the half of the wire surface which does not face the filament.

In the case of the anode also covering (the inside) of the anode with zirconium affects the characteristic. In *fig. 7* the i_a-V_a characteristics of two 1 000 W transmitter tetrodes are shown. One has a bare molybdenum anode; in the other the anode is covered internally with zirconium. In the latter

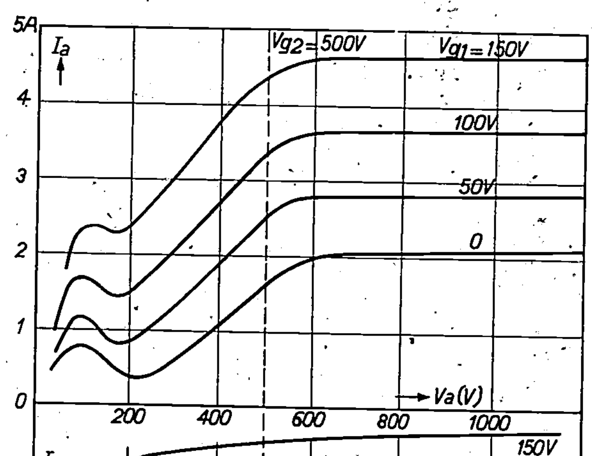


Fig. 7. The i_a-V_a characteristics of two tetrodes. The upper characteristic refers to a tetrode with molybdenum anode; the lower characteristic is obtained when the anode is covered internally with zirconium powder.

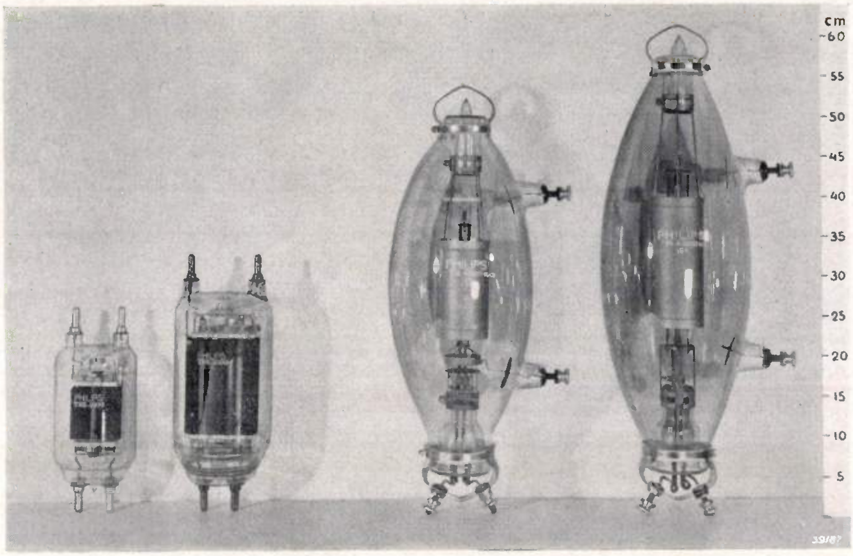


Fig. 8. Two modern transmitter triodes (left) and two used formerly (right) of about the same output. The photograph illustrates the enormous decrease in external dimensions.

case the familiar link which occurs as a result of the secondary emission for $V_a < V_{g_2}$ has disappeared almost entirely, and a characteristic is obtained which may compare with that of an ideally constructed pentode [5]).

Another favourable property of the powdered zirconium is its great capacity for radiating heat, which amounts to 80-90 per cent of that of a black body [6]). Because of this the specific dissipation of the anode and grids could be further considerably increased. It is clear from *fig. 8* how this permitted a reduction in dimensions compared with the earlier valves.

Thanks to the measures here described a great improvement could be obtained in the short wave properties. Most of these transmitting valves can be used on wave lengths down to 3 m without the efficiency, which amounts to 70 to 75 per cent for long waves, being too much diminished (see *fig. 9*).

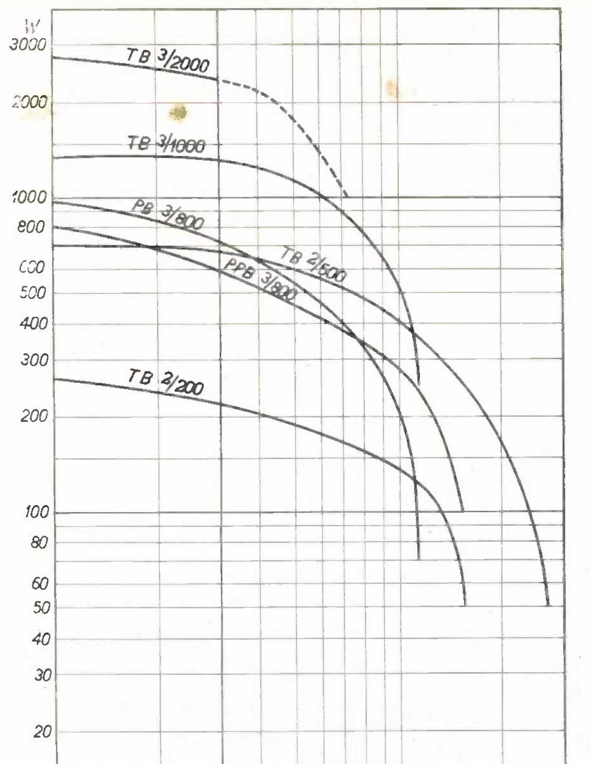


Fig. 9. Survey of the maximum obtainable outputs of the new series of transmitter valves on short waves. In different cases the triodes give a higher output than the corresponding pentodes; this takes place, however, at the expense of a much higher excitation power.

[5]) In judging this characteristic it must be kept in mind that different requirements are made of a transmitter valve than of a low-frequency output amplifier valve. Thus the "zirconium tetrode" when used as low-frequency amplifier will still give too much distortion, while in excitation or amplification of high-frequency energy this is of no importance.

[6]) Other finely divided substances such as tungsten also possess this property (see the article referred to in footnote [3]).

LAMPS FOR USE IN PHOTOGRAPHY

by N. A. HALBERTSMA and J. A. M. van LIEMPT.

771.448.1

In lamps for photographic purposes the temperature of the filament is much higher than in lamps for ordinary use at the expense of the life of the lamp. A much higher luminous efficiency is obtained, however, and this gain is manifested more strongly by the sensitive photographic plate than by the eye. Quantitative data are given on this subject. In conclusion the different types of electric lamps for photography by artificial light are briefly discussed.

At the present time there are many kinds of sources of light available for making photographic exposures by artificial light. Besides magnesium light, which was formerly much used for photography by artificial light, and from which the flashlight lamp was developed [1]), electric lamps are commonly used, especially in studios. They give of course less intense illumination and are also less easily moved about than the flashlight lamp, since this lamp carries its own store of energy, while the incandescent lamp must be connected to the light mains. On the other hand, however, it may be regarded as an advantage that with electric lamps the object to be photographed can be studied beforehand in the full illumination in which it is to be photographed, and that the expense of the electric lamp per exposure in less than that of a flashlight lamp. Whether the advantages of the flashlight lamp or those of the electric lamp are to predominate will depend on the purpose in view. So each of the two types of light sources has its own field of application.

In recent years electric incandescent lamps have been replaced in many fields of illumination by discharge lamps. Some of these discharge lamps, as *e.g.* the sodium lamp, exhibit unexpectedly favourable properties for photographic use [2]). The incandescent lamp, however, because of the ease of use and the absence of auxiliary apparatus, will maintain its importance in this field, so that it is justifiable to manufacture lamps adapted to the special needs of photography. The following account of the development of such lamps may serve to facilitate a choice among the different types of lamp which are manufactured for photographic uses.

Actinic radiation intensity and length of life

The choice of temperature for the filament of an electric lamp represents a compromise between two requirements: that of a high luminous efficiency and that of a reasonable length of life. It is possible to make lamps of very different life, but the longer the life the lower the temperature of the filament as well as the efficiency will be. Lamps for general lighting purposes are designed for an average life of 1000 hours. The luminous efficiency then amounts to 10—16 lm/W, depending upon the size and the voltage of the lamp. The permissible load on the house fuses or group fuses does not permit generally to connect more than 2 lamps of 500 W, corresponding to a light flux of 16,000 lumens.

Although the illumination which is thus obtained does not amount to more than a few per cent of average daylight, the adapted eye will consider a room illuminated with the light flux mentioned as "bright as daylight". For photographic materials, however, where the phenomenon of adaptation is absent, the flux of light mentioned will be insufficient in many cases, especially when, for example for the sake of depth of focus, the large apertures of modern types of lenses cannot be used.

It is therefore desirable to increase for photographic purposes the efficiency of electric incandescent lamps at the expense of the length of life. Since these lamps are only used for a few minutes at a time, their life is of only secondary importance and a considerable increase of efficiency is not too dearly obtained with the reduction of life to a few hours.

The increase of luminous efficiency, when the temperature of the filament is raised, may be explained as follows: in addition to an increase of the radiation for each wave length, the spectral composition of the radiation is shifted towards shorter wave lengths (see *fig. 1a*) and since the greatest part of the emitted radiation lies in the infrared, this means that a larger part of the radiation becomes visible. Since the maximum sensitivity of photographic emulsions lies at shorter wave lengths than that of the eye, the photographic efficiency of the radiation, when the temperature is raised, will increase more rapidly than the luminous flux. This photographic action, the ac-

---

[1]) For the flashlight lamp "Photoflux" see Philips techn. Rev. **1**, 289, 1936.

[2]) On the application of sodium lamps and other gas-discharge lamps in photography see Philips techn. Rev. **2**, 24, 1937; **3**, 91, 1938 and **4**, 27, 1939.

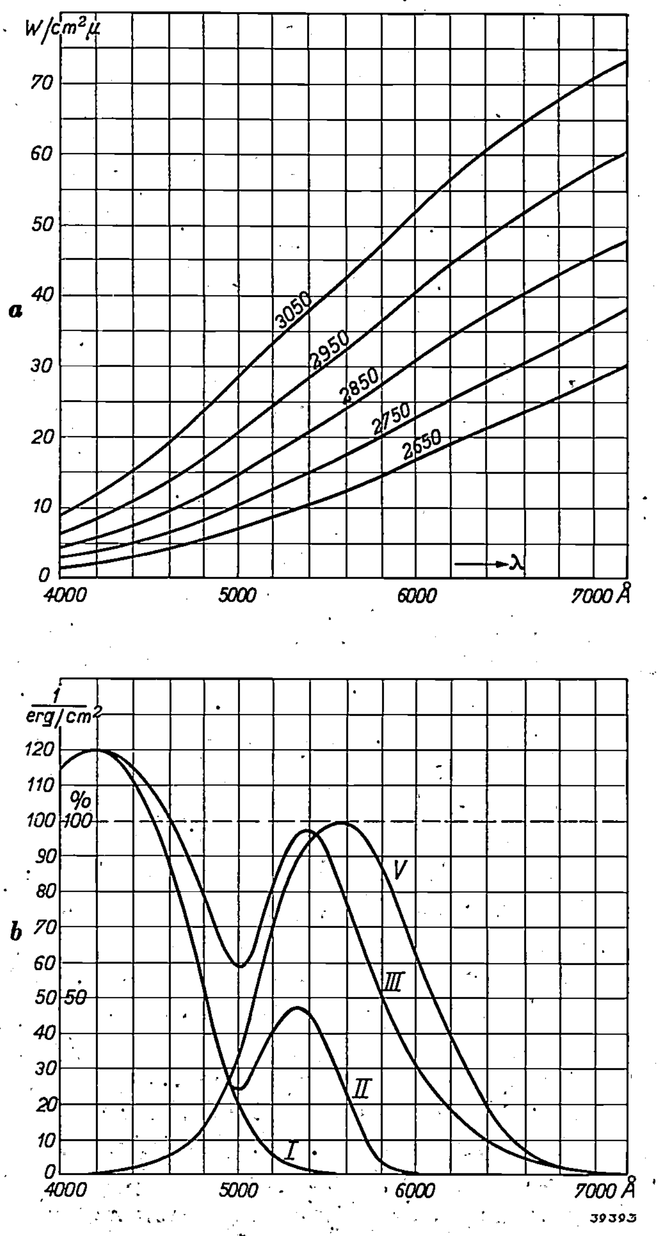


Fig. 1. *a*) Spectral intensity distribution of an incandescent tungsten surface for different temperatures which are indicated in °K.

b) Spectral sensitivity curve *V* of the eye and the sensitivity curves of different kinds of photographic emulsions (schematically): *I* non colour-sensitized, *II* orthochromatic, *III* panchromatic.

As a measure of the photographic sensitivity the reciprocal of the value of the radiation energy in erg/cm² which is necessary to obtain a density 0.1 above fog is used.

tinic value of the radiation, is of course also dependent on the nature of the light-sensitive emulsion; in the case of panchromatic emulsions the actinic value of the radiation will be much higher, and its variation with temperature will be found to correspond more closely with that of the visual efficiency than in the case of orthochromatic or ordinary, non colour-sensitized emulsions.

As an example, the average spectral sensitivity curves of a panchromatic, an orthochromatic and an ordinary, non colour-sensitized emulsion are reproduced in fig. 1*b* together with the relative eye-sensitivity curve *V*. With the help of the radiation curves of fig. 1*a* the visual and actinic efficiencies of the radiation have been calculated from these curves by a numerical integration for each temperature [3]).

The results are given in *table I*. Besides the visual intensity (light flux and three different actinic intensity values of the radiation), the table also gives the voltage, the power consumed, the life and the luminous efficiency for each temperature indicated. Since we are only interested in relative values, the values in all the columns were set equal to 100 for the normal filament temperature of 2650 °K, with the exception of the length of life, for which 1200 hours were taken [4]).

The table shows that the visual light intensity and the actinic value, like the length of life, vary very much with the temperature. If the temperature of the filament is raised by one per cent, the visual light intensity of the radiation increases by an average of 9 per cent. The actinic radiation intensities, on the other hand, rise by from 10 to 12 per cent, so that the actinic effect per unit of light flux becomes greater with increasing temperature. On the other hand, a temperature increase of one per cent causes the life to decrease by not less than 30 per cent. From this is follows that in order to achieve a reasonable improvement in actinic intensity a very large reduction of lamplife must be accepted.

The different lamps for photographic purposes

Philips lamps for photography may be divided into two classes: lamps with a life of 100 hours and lamps with a life of 2 hours. For general photographic purposes lamps with a life of 100 hours will generally be used; lamps of 2 hours life should, however, be preferred when very high intensities of illumination are required, as in making instantaneous exposures and cinematographic exposures or in working with colour film. Amateur photographers, as a rule, use these 2-hour lamps.

---

[3]) The possibility of calculating the sensitivity for heterochromatic light, starting with the sensitivity of the eye or of a photographic plate for light of different wave lengths, is based upon the addition law, see Philips techn. Rev. 6, 1941. The fact that the addition law is really valid for photographic emulsions has been demonstrated by A. v. Kreveld, Diss. Utrecht 1933 and Physica 1, 60, 1933.

[4]) The length of life depends not only on the filament temperature, but also upon the dimensions and the form of the incandescent filament. Therefore no generally valid conclusions can be drawn from the values in the table. The dimensions of the filament here assumed correspond approximately to those of the "Photolita" lamp, type S.

Table I

Properties of an electric lamp with tungsten filament for different temperatures *). The data for the calculation are due to C. Zwikker, Physica 5, 252, 1925.

Temperature	Colour temperature	Voltage	Power consumed	Life	Light flux	Luminous efficiency	Actinic value of radiation (%)		
°K	°K	%	%	h	%	%	panchromatic	orthochromatic	non colour-sensitized
2 650	2 710	100	100	1 200	100	100	100	100	100
2 750	2 820	113.5	119	270	141	116	148	154	159
2 850	2 930	125.2	141	70	196	135	212	230	244
2 950	3 040	138.4	164	19	265	156	299	336	365
3 050	3 150	153	191	6	349	181	408	473	526
3 150	3 260	169	223	2	455	209	548	657	747
Average percentage increase of above quantities for 1 per cent temperature increase									
1	0.94	2.89	4.64	—31	9.0	4.3	10	11	12

*) The actinic values for the panchromatic, the orthochromatic and the non-sensitized emulsions are set equal to 100 for a temperature of 2 650 °K, and may not therefore be compared with each other. To obtain comparable figures the values for the orthochromatic and panchromatic emulsions should be multiplied by 1.57 and 3.56, respectively.

Besides the length of life of the lamp, its power consumption and light distribution are also features to be considered.

The lamps are made in sizes of 500 watts and 250 watts, for all the usual mains voltages. With a life of 100 hours lamps of 500 watts have a light flux of 11 000 lumens, while with a life of 2 hours they develop a light flux of 16 000 lumens. The latter flux is about twice as large as that which an electric lamp of normal life would radiate for the same power.

The gain actually obtained, however, is, as already stated, greater than represented by the ratio of 1 : 2, since the actinic value of the light has increased considerably. If the average of the examples given in table I (the orthochromatic type of plate) is taken as a basis for the comparison, a lamp for photographic purposes of 500 watts, 11 000 lumens has the same actinic value as a regular electric lamp of about 13 000 lumens burning at normal temperature, while the photographer's lamp of 500 W, 16 000 lumens has an actinic value equal to about 25 000 lumens of ordinary electric light. If ordinary coiled-coil lamps would be used for photography, the largest type of which has an efficiency of about 15 lumens/watt, it would be necessary to install a total of 870 watts instead of one 500 W lamp of 100 hours and of 1 670 watts instead of one 500 W lamp of 2 hours life. In the case of the lamp with the life of two hours the power consumed will therefore be reduced in the ratio 3 : 1, which means, that with a given permissible load on the mains an actinic light flux can be excited with the lamp of 2 hours which is three times as large as that obtained by coiled-coil lamps. Colours are also reproduced more faithfully by these lamps than by daylight.

The light distribution of the different types of lamps depends upon the shape of the bulb, which may be partially covered with a mirror and on the finish of the bulb (clear, frosted, opal). A certain variety of types enables the photographer to con-

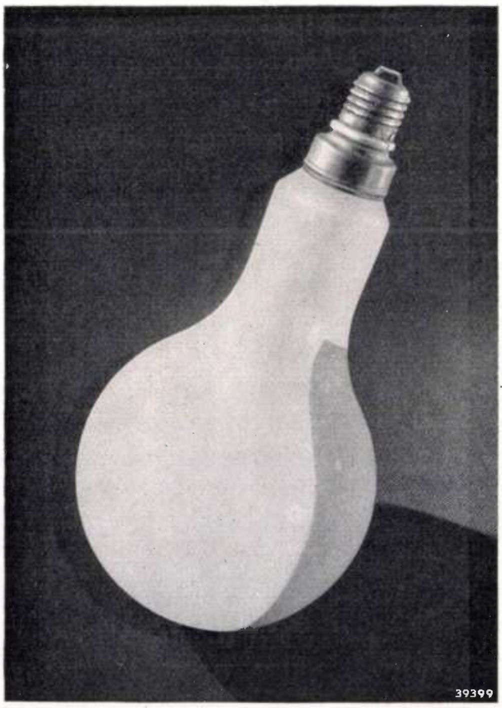


Fig. 2. The "Photomirenta" lamp with opal-glass bulb, one side of which is covered internally with a mirror.

centrate the light in various degrees depending upon the object to be photographed. By this a further appreciable gain can be achieved in the illumination of the object. We shall now briefly discuss the various types separately.

Lamps with a life of 100 working hours

The lamps with 100 working hours are manufactured in 2 types: the "Photomirenta" lamp and the "Argaphoto" lamp. Both have 500 watts, 11 000 lumens, and thus, according to the above, an average actinic intensity of 13 000 lumens. The "Photomirenta" lamp (see *fig. 2*) has a spherical opal-glass bulb 15 cm in diameter, one half of which is covered internally with a mirror. In this way a wide beam is obtained which by the diffusing effect of the opal-

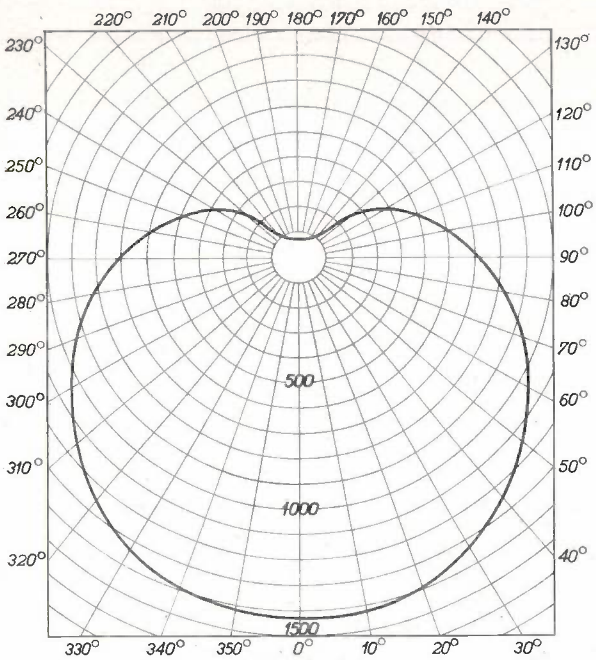


Fig. 3. Light distribution curve of the "Photomirenta" lamp (see fig. 2) in a plane perpendicular to its axis.

glass, gives no harsh shadows. In *fig. 3* the light distribution is shown. This light distribution forms a satisfactory compromise for universal application in the studio of the professional photographer.

The "Argaphoto" lamp has a small bulb without mirror, so it should be used in combination with a reflector. In *fig. 4* the light distribution curve of a "Philiray" reflector SC 255 with "Argaphoto" lamp is shown. The maximum intensity is 20 times the mean spherical candle-power of the lamp.

The "Argaphoto" lamp with reflector is especially suitable for taking photographs with artistic effects, or for providing additional illumination to that obtained with the "Photomirenta" lamp. For infrared photography also, with or without an infrared filter, it is a very satisfactory source of light.

The lamp is furthermore important in taking photographs with less sensitive kinds of films, such as the so-called duplicate film (which can be developed directly into a positive), and the modern colour films for miniature cameras.

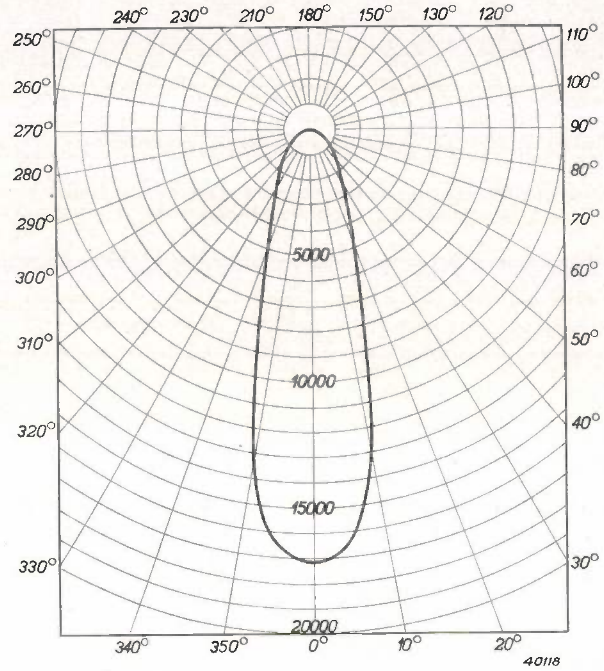


Fig. 4. Light distribution curve of the "Argaphoto" lamp with "Philiray" reflector, type SC 255.

Lamps with a life of two hours

The lamps of two hours are made in two sizes: 250 watts with a flux of 9 000 lumens, and 500 watts with a flux of 16 000 lumens.

These "Photolita" lamps, type SM and NM,

Fig. 5. "Photolita" lamp with mirrored bulb in an adjustable lampholder.

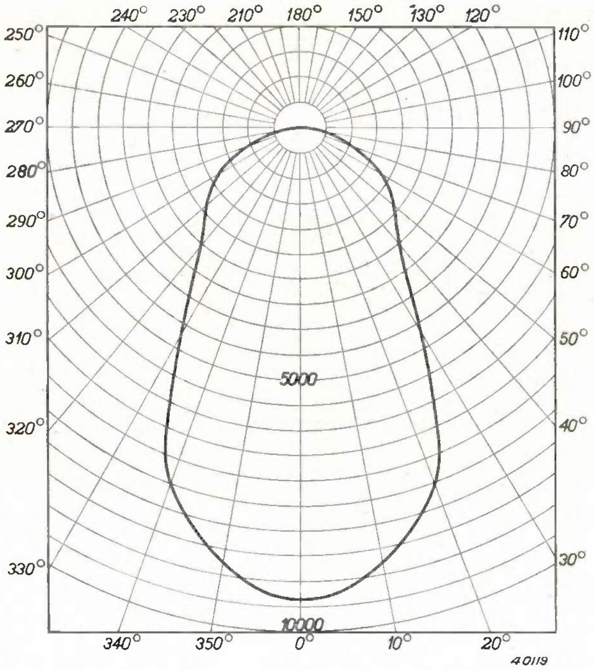


Fig. 6. Light distribution curve of the "Photolita" lamp with mirrored bulb.

have a bulb of a special shape, part of which is provided with a specular reflecting layer. The outer surface of the bulb is frosted. These lamps can therefore be used without reflector, which greatly facilitates their use by amateur photographers (see *fig. 5*). "Photolita" lamps type S (250 watts) and N (500 watts) are made without reflecting layer, the bulb being inside frosted. They should only be used with a separate reflector. *Fig. 6* presents the light distribution curve of the "Photolita" lamp type NM.

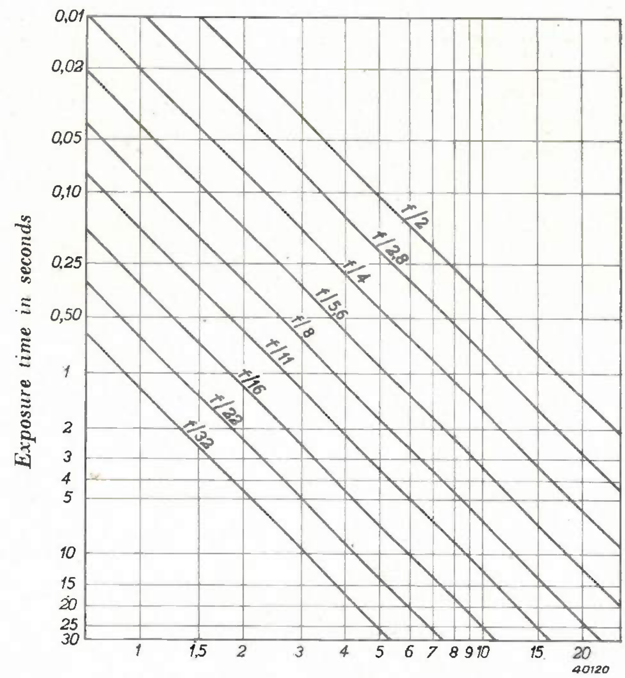


Fig. 7. Exposure diagram for a "Photolita" lamp NM or N (the latter being used in a mirrored reflector. The diagram is based on the use of orthochromatic material of 4 000 H. & D. or of panchromatic material of 2 400 H. & D.

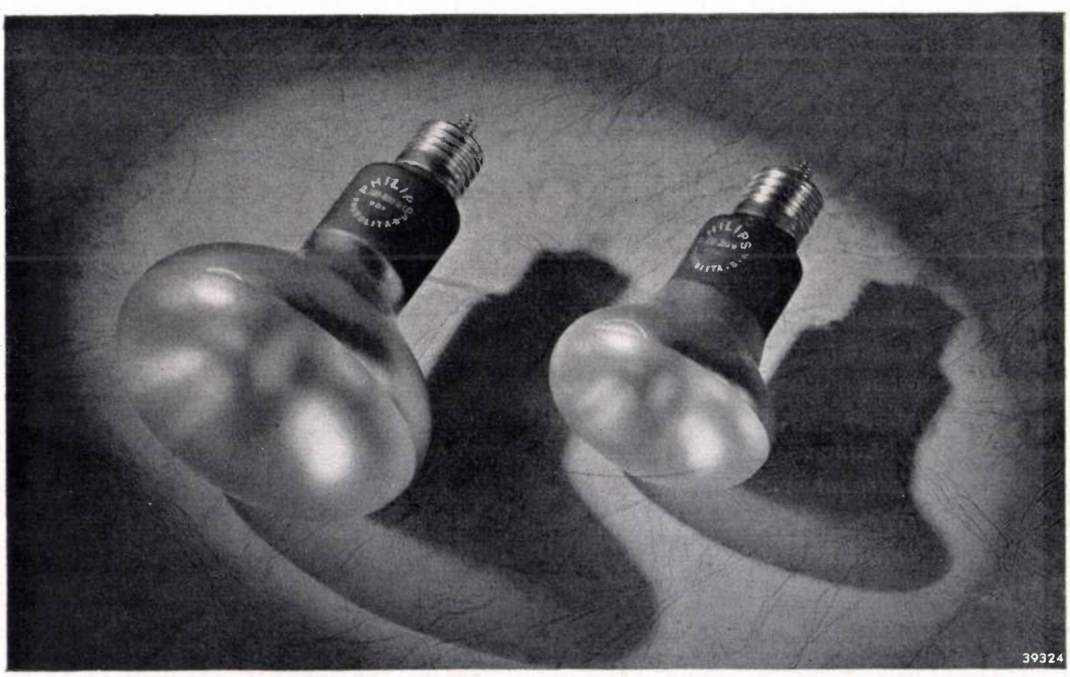


Fig. 8. The "Photolita" lamp with internal mirror. In order to prevent radiation of light toward the rear, the neck of the bulb is blackened. The base of the lamp is provided with a small spiral spring in order to make it possible to screw the lamp into the lampholder in the prescribed position.

In connection with the highly loaded filament some precautions should be taken in using the "Photolita" lamps in order not to shorten still more the already short life. Lamps with mirror must not burn in any position, but only in that indicated on the bulb of the lamp. Moreover, the mains voltage must never be higher than that marked on the lamp. Care must also be taken not to switch on the lamp for a longer time than is absolutely necessary for making the exposure. A convenient method of accomplishing this is the use of a series-parallel switch with which two lamps can be connected in series or in parallel as desired. For the arranging of the light sources, which in careful work may take considerable time, the lamps are connected in series, so that they burn at reduced voltage and have a practically unlimited life (about 1 000 hours). Shortly before the exposure the lamps are connected in parallel. The exposure diagram of *fig.* 7 shows the exposure times which can be applied, when using "Photolita" lamps.

Lamps for exact colour rendering

The combination of any given light source with any kind of negative material does not in general produce a correct colour rendering, *i.e.* the blackening of the negative does not entirely correspond to the brightness values of the different colours by daylight. If an orthopanchromatic or an panchromatic emulsion is used for the negative these inaccuracies will usually not be disturbing when electric lamps are used. For very high requirements, the colour rendering can sometimes be slightly improved by the use of a blue-green or green filter.

For still more red-sensitive pan-material special lamps are made provided with an inside-frosted blue-glass bulb. There are two sizes, the photographic lamps type C of 500 watts and type E of 1 000 watts. The life of these lamps is 300 hours. These lamps can be used for extra illumination in daytime, when making colour photographs on daylight film, which does not give a correct colour rendering with the light of ordinary electric lamps without filter.

A SIMPLE SYSTEM OF BAND SPREAD IN SHORT-WAVE RECEPTION

by C. J. van LOON. 621.396.662 : 621.396.62.029.58

While in the system of band spread described previously in this periodical use was made of separate variation elements (variable condenser or coil) for the accurate tuning within each of the five or six short-wave bands, a system has now been worked out in which the ordinary rotary condenser used for normal tuning is also used for this accurate tuning. The range of variation of the circuit capacity is in this case reduced to the desired small size by the additional connection of fixed condensers. The results of this, and several particulars of the practical construction are discussed. The most important advantages of the system, in addition to a better signal-to-noise ratio and complete freedom from microphonic effect, are the simplicity of the mechanical construction and of the operation. The latter is now reduced to the same manipulation as the tuning to long-wave and intermediate-wave stations and also takes place with the same knobs.

Introduction

The resonance frequency of an oscillation circuit which is composed of a fixed self-induction coil and a rotary condenser of the usual construction generally changes between the extreme positions of the rotary condenser approximately by a factor 3.5. For the tuning of radio receiving sets, therefore, the region of the long waves (about 700-2000 m), the intermediate-wave region (60-560 m) and the short-wave region (13-50 m) can each be covered with the help of one definite coil which is connected to the oscillation circuit by means of the so-called wave-length switch.

In the region of short waves from 13 to 15 m a complete turn of the rotary condenser thus corresponds to a frequency variation of about 22 500 — 5 800 = 16 200 kc/s. If two transmitting stations in this wave-length region lie side by side with the customary interval of 10 kc/s between them, an extremely slight twist of the rotary condenser (usually 1/1 600 of the full turn) will change the tuning from the one transmitter to the other. In order to make the tuning just as easy at these short wave lengths as we are accustomed to with ordinary broadcasting waves, there would have to be an extremely fine and accurately reproducible setting of the rotary condenser as well as a very much elongated station dial.

The structural difficulties which are hereby encountered have led to a search for a solution in a different direction. This solution is based upon the fact that the broadcasting stations in the region of short waves mentioned are concentrated according to the present international agreement in only seven relatively narrow bands, see *table I*.

Table I

Band	Frequency limits in megacycles/sec	Δf in kc/s	$\Delta f/f$
13 m	21.45 - 21.75	300	1.4%
16 m	17.75 - 17.85	100	0.6%
20 m	15.10 - 15.35	250	1.7%
25 m	11.70 - 11.90	200	1.7%
30 m	9.50 - 9.70	200	2.1%
40 m	7.20 - 7.30	100	1.4%
50 m	6.00 - 6.20	200	3.3%

Each of these bands may be considered as a separate wave region analogous to the regions of long waves and intermediate waves, but with the difference that within each of the bands the frequency need not be varied by a factor 3.5 but only by a few per cent (column $\Delta f/f$ table I). Therefore one may tune to every band with the help of an absolutely constant, fixed coil and condenser, and the required slight frequency variation within the band can be obtained with an extra variable coil or condenser whose self-induction or capacity need only be a few per cent of that of the fixed elements, and therefore need not be so accurately reproducible. Each band is in this way spread over a full turn of the variation element and over a corresponding separate frequency scale with the normal density of stations. This is the origin of the common name of "band spread" for these tuning systems.

In a previous article in this periodical [1]) the principle and two different systems of band spread in practical use have already been discussed. In the case of the first system the tuning to each band was obtained by combination of certain fixed coils and condensers. The set in which this system was used was a superheterodyne receiver with three tuned circuits, namely aerial circuit, high-frequency amplifier circuit and oscillator circuit; the band spread was, however, applied only to the last cir-

[1]) Radio sets with station dials calibrated for short waves, Philips techn. Rev. 4, 284, 1939.

cuit. In order to tune the other two circuits which work without band spread at least roughly, they first had to be set on the band with the ordinary set of rotary condensers; then by means of a switch the condenser group of the oscillator circuit was replaced by a fixed capacity with a separate small variable condenser in parallel with it. The latter had a separate drive and station dial by means of which it was now possible to tune accurately within the band.

In the case of the second practically constructed system the same fixed coil was used for all the bands and the tuning to each band was accomplished by putting the set of variable condensers (linear action condensers) used for the ordinary tuning into a definite very accurately determined position mechanically. Thus all three circuits were already tuned to the band, and in two circuits, namely that of the oscillator and that of the high-frequency amplifier stage, band spread was now applied by tuning within each band with a small variable self-induction (coil with sliding iron core), which again had its own drive and dial.

We shall not go into the specific advantages and disadvantages of these systems; attention must only be called to the fact that the introduction of each of these band spread systems caused the appearance of two new operating knobs on the front panel of the set, namely one for the choice of the band and one for tuning within the band. Although of itself this is contrary to the universal attempt to keep the operation of receiving sets as simple as possible, it was accepted in the expectation that when new facilities are offered the user will accept a somewhat more elaborate operation into the bargain. Then after some time, when the novelty has worn off, in addition to improvements the requirement of simple operation again becomes important, while at the same time the manufacturer will also try to use the new invention in cheaper sets by simplifications in the construction, and thus make it available to a larger circle of users.

In conformance with this general line of development, a new system of band spread has been worked out by Philips, which makes the operation just as simple as it formerly was. The basic idea is that it represents a certain extravagance to use a separate variation element for the tuning within the band while there is already a variable element present in the shape of the ordinary rotary condenser. The latter has of course a range of variation of the capacity which is several hundred times as large as is required for the tuning in each band. The variation of the circuit capacity resulting from this can, however, be reduced to the desired size in a relatively simple way. The separate operation arrangements for the band spread may then disappear again: the choice of band can be made with the same wave switch and the tuning within the band with the same tuning knob which is used for the intermediate and long-wave regions [2]).

The use of the ordinary tuning condenser as variation element

How is the reduction of the variation range accomplished? More precisely stated, how can it be brought about that upon a full turn of the rotary condenser the total capacity C of the oscillator circuit varies only by a small amount ΔC, and $\Delta f/f = {}^1/_2\, \Delta C/C$ takes on the values indicated in table I (or slightly larger)?

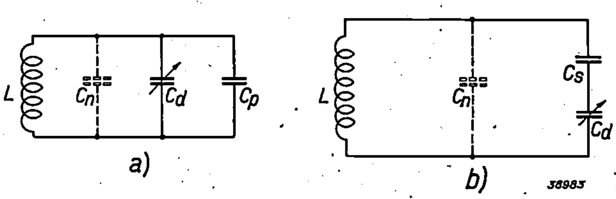


Fig. 1. In the oscillator circuit consisting of the coil L and the rotary condenser C_d a fixed condenser can be connected in parallel (C_p in a) or in series (C_s in b) with the rotary condenser. C_n represents the parasitic capacities in the circuit.

The simplest way of reducing the variation range of the capacity is to connect a fixed condenser in parallel or in series with the rotary condenser, see *fig. 1a* and *b*. Let us first consider fig. 1*a*. In this diagram C_n represents all the parasitic capacities of the circuit. Without C_p the circuit capacity would change from the small values $C_n + C_{d\,min}$ (when $C_{d\,min}$ is the zero capacity of the rotary condenser) to the value $C_n + C_{d\,max}$ upon a complete turn of the rotary condenser. In a practical case the following may be true:

$$C_n = 30\ \mu\mu\text{F},$$
$$C_{dmin} = 10\ \mu\mu\text{F},$$
$$C_{dmax} = 450\ \mu\mu\text{F}$$

so that the circuit frequency may vary by the factor $\sqrt{450/40} \approx 3.5$ already mentioned. After the connection of C_p, however, the ratio between maximum and minimum capacity becomes

$$\frac{C_p + C_n + C_{dmax}}{C_p + C_n + C_{dmin}}$$

and it is clear that when the capacity C_p in parallel

[2]) In the first practical model of the new system, in order not to make the wave switch too elaborate or complicated, an extra switch was introduced for the choice of the band. This is, however, not essential and can be avoided by using a different switch mechanism.

is chosen large enough the frequency variation $\Delta f/f$ upon a full turn of the rotary condenser can be reduced to any desired small value. For example, in order to make $\Delta f/f = 2$ per cent in the foregoing specific example C_p would have to be 11 000 pF. If fig. 1*b* is considered in the same way it is clear that the same result can be obtained by the connection in series of a sufficiently small capacity C_s; in the example for $\Delta f/f = 2$ per cent the series capacity C_s would have to be about 4.2 pF.

In the first case therefore the circuit capacity is made very large, in the second very small. Neither of the two cases therefore is of direct practical use. The circuit capacity may not be too small, since then the parasitic capacities of coil windings and connections, which usually depend very much on the temperature and other influences, make up too large a part of the whole capacity and their variations would therefore cause the circuit capacity to vary too much. On the other hand, especially when we consider the oscillator circuit of a superheterodyne set, the circuit capacity may not be too large, since it would then become difficult to cause the circuit to oscillate. The diagram of the oscillator circuit shown in *fig. 2* illustrates this. At a certain grid A.C. voltage on the valve an anode

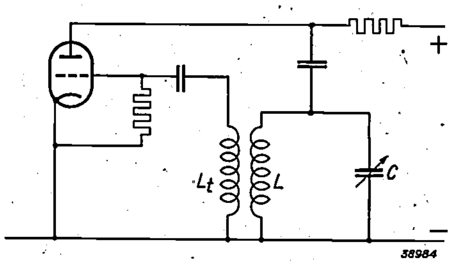


Fig. 2. Diagram showing the principle of the connections of the oscillator of a superheterodyne receiving set.

alternating current flows. This causes an A.C. voltage on the LC circuit and this in turn an A.C. voltage in the back-coupling coil. This A.C. voltage must be (at least) equal to the original grid A.C. voltage in order that the oscillation shall not die out. Now this A.C. voltage will be greater the tighter the coupling of L and L_t, and the larger the impedance in parallel with the $L.C.$ circuit (*i.e.* the voltage on L). This impedance is given for the resonance frequency ω by $R = Q/\omega C$, where Q is the quality factor of the coil. To make the circuit oscillate, therefore, especially on short waves, the back-coupling must be tighter the greater the capacity C. Structurally it is not so easy to realize the desired tight back-coupling, especially since a smaller coil L is again necessary for the tuning with larger C.

These considerations led to the requirement for the short waves here considered that the circuit capacity must not be smaller than about 150 pF or larger than about 250 pF.

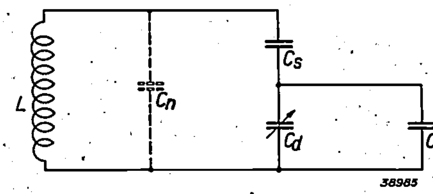


Fig. 3. Connections with which the percentage variation of the circuit capacity upon a full turn of the rotary condenser C_d can be reduced to any desired small value, while the total circuit capacity can take on a prescribed value.

This requirement, together with that of a given value of $\Delta C/C$, can be satified by connecting two fixed condensers C_p and C_s as indicated in *fig. 3* to the rotary condenser. With this complex of capacities the total circuit capacity for the minimum and maximum positions of the rotary condenser amounts to

$$\left.\begin{aligned} C_1 &= C_n + \frac{1}{\dfrac{1}{C_s} + \dfrac{1}{C_{dmin} + C_p}}, \\ \text{and}\quad C_2 &= C_n + \frac{1}{\dfrac{1}{C_s} + \dfrac{1}{C_{dmax} + C_p}}, \text{ respectively.} \end{aligned}\right\} \quad (1)$$

We may now prescribe definite values for C_1 and C_2, *i.e.* for the circuit capacity and its variation, and we then have in (1) two equations from which the desired values of C_s and C_p can be calculated. If for example with the values of C_n, C_{dmin} and C_{dmax} we wish to obtain the values $C_1 = 162$ μμF, $C_2 = 171$ pF, we find that we must choose; $C_s = 160$ pF and $C_p = 750$ pF.

If the aim is to spread the frequency region Δf of each of the bands of table I (plus a small frequency region as reserve on either side) over the complete deviation of the rotary condenser, a different C_1 and /or C_2 must actually be prescribed for each band, since according to table I, $\Delta f/f$ which is equal to $^1/_2\ (C_1 - C_2)/C_1$ is smaller for the short waves than for the long. Now in order not to make the switch which serves for the transition from one band to the other too complicated, the same combination C_s, C_p is used for all bands, while a fairly equal spreading of the different bands is obtained by connecting some extra capacity in parallel with the whole for shorter waves (*i.e.* C_n in equation

(1) is increased). In this way the largest circuit capacity is obtained at the shortest wave length. This is also an advantage in connection with the drift of the frequency: it is exactly at the shortest waves, where the tuning is most sensitive to small capacity variations, that the parasitic capacities subject to drift have the least influence.

Even without the desirability of equal spread a small adjustment capacity (trimmer condenser) must be added for each band in parallel with C_n, in order to be able to tune the circuit accurately to the band, even with slight deviations in the capacity and self-induction values of the fixed condensers and coils used. It is of course a primary requirement that the capacity and self-induction values mentioned should drift so little when in use that after the adjustment in the factory the trimmer condenser need not be set again.

Incidentally, it may be stated that it was also desirable to leave unchanged the combination of C_s, C_p for all bands, in order to obtain the same scales for the bands for every set in series manufacture. The tuning, especially on high frequencies, is very much affected by the parasitic self-inductions of the connections between C_s, C_d and C_p, which are subject to variations. A deviation in length of 1 mm of the connecting wires (corresponding to a deviation of the self-induction of 0.001 μH) already changes the calibration of the scale for the 13 m band by 15 kc/s! Due to the fact that with fixed values of C_s and C_p the connection lines could be kept very short, it was possible to make the parasitic self-inductions sufficiently reproducible.

Other particulars of the connections

With fixed C_s, C_p the wave switch with which the bands are selected need only provide for the inclusion in the connections of the correct coil and of the supplementary capacities mentioned for each band. Since with an increasing number of positions of the switch the difficulties in its construction increase rapidly, we limited ourselves to spreading five of the bands in table I. We chose the first five since in the case of the last two it is relatively less difficult to tune in the ordinary way. Even with this limitation to five bands there occurred a difficulty, although it was not serious, in connection with finding space for the required fixed coils. The coils were mounted in standardized cans each of which can hold two coils. Each additional can means an increase of the space necessary on the chassis and in general an enlargement of the whole set. In order to keep the set as small as possible it was very desirable to use as few coils as possible. Now by choosing C_s and C_p within the available limits, so that the same coil could be used for the 30 m band as well as for the continuous tuning from 13 to 50 m, and also by using a connection in parallel of coils for the other bands, the number of coils for the whole short-wave region could be reduced to four (*i.e.* two cans).

Until now we have always been speaking of one circuit, namely the oscillator circuit, which is of primary importance for the band spread, we saw that in the system of band spread here considered the ordinary rotary condenser of the oscillator circuit is used for the tuning within the bands. Now the rotary condensers of the two other circuits (aerial circuit, high-frequency amplifier circuit) which serve for the ordinary tuning are, however, mounted on the same shaft as the rotary condenser of the oscillator circuit. Since the other circuits must also be tuned at least roughly to the bands, and since the ordinary rotary condensers can no longer be used for this, we must now provide separate tuning elements for these circuits.

This apparent complication is automatically solved when we apply band spread to the other two circuits as well. Tuning to the bands then again takes place by the switching in of fixed coils, and tuning within the band by the ordinary rotary condensers, in parallel and in series with which fixed condensers C_p and C_s, respectively, are connected in exactly the same way as in the oscillator circuit. The three switches which serve for the selection of the band in the three circuits (connection of coils, supplementary condensers, etc.), are placed on the same shaft, like the three rotary condensers. In the end, while the band spread is still better than in the earlier systems described in the article cited [1]) (it is now applied to all three circuits), the operation is nevertheless simpler, the same tuning manipulations are required as in the long-wave and intermediate-wave regions.

The application of band spread in all three circuits involves the use of a larger number of fixed coils and condensers than in the case of band spread in the oscillator circuit alone. There are, however, other important advantages besides the simplicity of operation. Without band spread, thus when the whole short-wave region must be covered with one turn of the rotary condenser, the circuit capacity must increase with the square of the wave length. The result is that, according to the above-mentioned formula $R = Q/\omega C$, there is a lower impedance in parallel with the oscillator circuit at longer waves and thus also less building up of oscillation (in the aerial circuit) or amplification (in the high-frequency amplifier stage). With band spread, however, where a different coil is used for every band, the circuit capacity can be chosen equally small in all the bands, or on longer waves (30 m) it may even be chosen smaller than in the other bands (see above), so that now a greater amplification and thus a better signal-to-noise ratio can be obtained over the whole region. This advantage is even more strongly emphasized by the fact that in the adjustment of the first two circuits such a great accurary is not required, and therefore such great care need not be taken to prevent a drift of the total circuit capacity. Because of this, with a given value of the parasitic capacities a smaller total circuit capacity is sufficient. In our case the fixed capacities C_p and C_s — in each circuit there is again only one set C_p, C_s for all five bands — have such dimensions that a total circuit capacity of about 80 pF was obtained, while without band spread the circuit capacity in the 30 m band would have had to be about 180 pF.

Summarizing it may be stated that there are numerous advantages in the system of band spread here described, although a fairly large number of fixed coils and condensers must be used. Not only is the tuning entirely insensitive to mechanical vibrations and shocks (absence of the microphonic effect) as in the case of all band spread systems with fixed condensers for the bands, but, moreover, the operation is just as simple as for the tuning on long-wave and intermediate-wave stations; furthermore no special expensive mechanical arrangements are necessary for the band spread, and the signal-to-noise ratio could be made considerably better than in the case of short-wave reception without band spread.

THE MAXIMUM ELECTRICAL FIELD STRENGTH FOR SEVERAL SIMPLE ELECTRODE CONFIGURATIONS [1])

by A. BOUWERS and P. G. CATH.

537.212 : 621.3.027.7

The maximum voltages which may occur on parts of high-voltage apparatus determine the chance of breakdown, either toward earth or toward other points. Formulae and tables are given for different configurations which may be of service in designing high-voltage apparatus.

Introduction

Apparatus and installations have repeatedly been described in this periodical in which very high electrical voltages occur, such as X-ray tubes and generators for very high direct or surge voltages [2]). In the construction of such apparatus one object is to decrease the dimensions still further while retaining the same properties. The occurrence of breakdown, which will usually be in air, when the electrical field strength at some point on the surface of one of the electrodes exceeds 30 kV/cm, determines in general how far this object may be pursued.

The smaller the dimensions of such constructions the more carefully the shape of the different components must be considered. The familiar fact that sparks jump much more easily between electrodes with small radii of curvature (points) than between electrodes with a large radius of curvature, with the plane surface as a limiting case, furnishes the simplest contribution to the knowledge which is required in all these cases.

In the articles referred to a number of methods are also discussed which may be used to avoid breakdown. For example, plane terminal electrodes are fixed to points with great curvature, and in this way the maximum field strength is decreased. For such high-voltage installations thin wires with a small radius of curvature are not used for the current supply, but thick hollow tubes which have a much lower field strength on the surface at the same potential. Components which are too sharply curved at certain points are entirely surrounded by spheres which are connected conductively to them. Since no electric field exists within the sphere, and the field strength on the outside of the enveloping sphere is lower on an average than at the sharp points of the components enveloped, breakdown takes place only at a higher potential.

In such considerations it is naturally important to be able to estimate the magnitude of the field strength in the dielectric with a chosen model. The means of doing this are in the form of calculations which are based fundamentally on Laplace's equation and on Gauss' theorem and on several simple empirical rules of calculation derived from the former. Together these allow the constructor to make a fairly sure estimate of the load on the dielectric in his design, which estimate is in this case of just as great importance as the insight into its mechanical strength. Great accuracy is here of less importance than the possibility of being able to apply the prescriptions efficiently, quickly and reliably to all the configurations which he encounters.

What electrode configurations are used in practical cases?

This question cannot be answered for all cases which occur, since each new construction may lead to different designs which have not formerly been used. If, however, we consider a fairly complicated apparatus such as the generator for 1.4 MV reproduced in *fig. 1*, which is intended for the Laboratoire de Chimie Physique in Paris, it strikes one that certain simple arrangements of spheres and cylinders are repeated over and over again. In other cases flat planes are encountered in combination with spheres and cylinders. This is not of course an arbitrarily accepted principle of construction, but the consequence of the necessity of keeping the field strength on the electrodes as small as possible with given potential differences and distances between the electrodes.

At a given distance and a given potential difference the maximum field strength on the electrodes is smallest when the field is homogeneous and the lines of force are thus parallel, as is the case between very large, charged, plane plates. Such plates, however, always have edges at which the field strength is locally much greater: it is greater the sharper the angle at which the surface passes from one direction to another. By rounding off edges, therefore, the transitions are made as gradual as possible. Curved surfaces will therefore be encountered in the construction; if they are curved in one direction cylindrical electrodes are obtained, if in two direc-

[1]) Adapted from Chapter II of A. Bouwers, Elektrische Höchstspannungen, Springer, Berlin 1940.

[2]) Philips techn. Rev. 1, 6 and 235, 1936; 2, 161, 1937; 3, 259, 292, 306, 331, 1938; 4, 153, 1939; 6, 46, 1941.

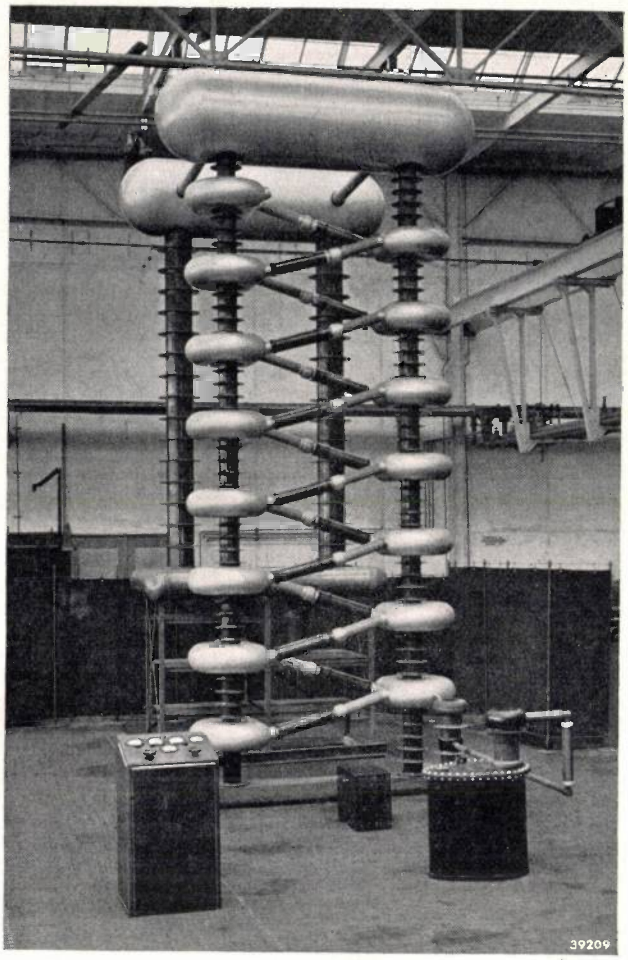


Fig. 1. Generator for 1.4 MV constructed for the Laboratoire de Chimie Physique in Paris, which exhibits the arrangement of doubly and singly curved surfaces often occurring in high-voltage constructions.

tions spherical or ellipsoidal electrodes (mushrooms). The simplest general cases with which we are concerned are therefore the following:

a) plane plates, no curvature,
b) cylindrical electrodes, curvature in one direction,
c) spherical electrodes, curvature in two directions.

Field distribution for several simple cases

In a space between charged electrodes of any given form the variation of the potential φ can in principle be calculated by solving the differential equations of the electrostatic field. This description of the electric field E (E is a vector, the component in a direction s is $E_s = -\partial\varphi/\partial s$) involves the fact that usable solutions must have a mathematical form which gives a constant value of the potential on the electrodes. For electrodes of any given shape it is difficult to satisfy this condition mathematically; in general it is easier to find a solution of a differential equation than to satisfy fairly complicated boundary conditions.

Even for very simple surfaces: plane plate, sphere or cylinder, elementary considerations are only possible when the second electrode is also a plane plate parallel to the first, a concentric sphere or a cylinder with the same axis, respectively. It is already difficult to calculate the field distribution between two charged plane plates which are not parallel. For two charged spheres which are not concentric and one of which surrounds the other the calculation of the field distribution is also not very simple. Fortunately the cases: parallel plane plates, concentric spheres and coaxial cylinders, are so important practically that we can carry out the main part of our task. We shall begin by giving briefly the results confining ourselves to the field in air (see later for other dielectrics).

1) With two parallel plane plates at a distance a from each other the potential φ which satisfies Laplace's equation

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

and the boundary conditions is:

$$\varphi = Ax + B,$$

where x is measured according to the normal and A and B are constants.

If $\varphi = \varphi_1$ for $x = 0$ and $\varphi = \varphi_2$ for $x = a$ one finds that

$$A = \frac{\varphi_2 - \varphi_1}{a} = -\frac{U}{a},$$

where $U = \varphi_1 - \varphi_2$. The field strength is

$$E_x = -\frac{\partial\varphi}{\partial x} = -A = \frac{U}{a}\cdot$$

It is constant and the field is thus homogeneous.

2) The field of a charged sphere (charge q) is easily calculated according to Gauss' theorem, which relates to a closed surface s which encloses a charge q, and is as follows:

$$\int E_n \,\mathrm{d}s = \int \frac{\partial\varphi}{\partial n}\,\mathrm{d}s = 4\pi q, \quad . \; . \; . \quad \text{(G)}$$

where n is the direction of the normal to the surface. With this one finds for the radial field

$$4\pi r^2 E_r = 4\pi q,\ E_r = \frac{q}{r^2}\cdot$$

3) In the case of two concentric spheres one finds from the foregoing that $\varphi = A/r + B$, where A and B are constants and r is measured from the common centre. If the radii of the two spheres

are r_1 and r_2 $(r_2 > r_1)$ and φ_1 and φ_2 are their potentials, then

$$\varphi_1 - \varphi_2 = A\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = U,$$

so that

$$E_r = -\frac{\partial\varphi}{\partial r} = \frac{U}{r^2}\frac{r_1 r_2}{r_2 - r_1}.$$

The greatest value of E_r is found at the surface of the inner sphere $(r = r_1)$. It amounts to:

$$E_1 = \frac{U}{r_1}\frac{r_2}{r_2 - r_1} = \frac{U}{a}\frac{r_2}{r_1}, \quad \dots \quad (1)$$

where $a = r_2 - r_1$ represents the distance between the surfaces of the two spheres. The maximum field strength is thus r_2/r_1 times as large as in the corresponding case of two parallel plates at a distance a apart. This result can also be represented by means of an efficiency factor η. This quantity is defined as the quotient of the electrical field strength E between two parallel plates at a distance a apart and the maximum electrical fieldstrength E_m in the case of non-planar conductors with a least distance a apart at the same potential difference U, so that

$$\eta = \frac{E \text{ (// plates; distance } a)}{E_m \text{ (conductors with least distance apart } a)}$$

In our case therefore

$$\eta = \frac{r_1}{r_2} = \frac{r_1}{r_1 + a} < 1.$$

The efficiency factor approaches unity when r_1 becomes infinite, a being kept constant, and it tends to zero as r_1 approaches zero. This value of η will be found in column (a) of *table I*. Its reciprocal $1/\eta$ (in our case r_2/r_1) indicates how many times greater the maximum field strength is than in the corresponding case of parallel plates at the same distance from each other.

The question may also be put as to how E_1 varies when r_2 and U are kept constant and r_1 is considered variable (a is of course not constant here). E_1 then has a minimum for $r_1 = {}^1/_2\, r_2$. The value of E_1 in that case is

$$\frac{2U}{r_1} = \frac{2U}{a},$$

while $\eta = 0.5$.

4) In the same way as in case 2) one finds for the field around a charged cylinder

$$2\pi r E_r = 4\pi q, \quad E_r = \frac{2q}{r},$$

where q represents the charge per cm length.

5) As in case 3) it may be derived from this for two coaxial cylinders $(r_2 > r_1)$ that

$$\varphi = A \ln r + B,$$

so that $\quad \varphi_1 - \varphi_2 = A \ln r_1/r_2 = U$

and $\quad E_r = -\frac{\partial\varphi}{\partial r} = \frac{U}{r}\frac{1}{\ln(r_2/r_1)}.$

The largest value of E_r again occurs at the surface of the inner cylinder. One finds there

$$E_1 = \frac{U}{r_1}\frac{1}{\ln(r_2/r_1)} \quad \dots\dots \quad (2)$$

For the efficiency factor one finds

$$\eta = \frac{r_1}{a}\ln\frac{r_1 + a}{r_1}.$$

Here also η approaches zero when r_1 approaches zero, a being constant, and η tends to unity as r_1 approaches infinity.

Furthermore one can again ascertain the value of r_1 for which the maximum field strength $E_m = E_1$ has its smallest value, with r_2 and U being kept constant. This is now the case for $r_2/r_1 = e = 2.718$. E_1 then has the value

$$\frac{U}{r_1} = 1.718\,\frac{U}{a},$$

while $\quad \eta = \frac{1}{1.718} = 0.58.$

Table I

Efficiency factor $\eta = E/E_m$ for fields on a sphere surrounded by a concentric sphere, compared with the same factor for a sphere and a plane plate and for two similar spheres at a distance from each other.

$\frac{r+a}{r}$	concentric spheres (a)	sphere and plane plate: rigorous (b)	$k = \frac{(a)}{(b)}$	$k = 0.9$ (c)	two similar spheres: rigorous (d)	$k = 0.9$ (e)
1.5	0.667	0.732	0.91	0.74	0.850	0.88
2	0.500	0.563	0.89	0.555	0.732	0.74
3	0.333	0.372	0.90	0.37	0.563	0.575
4	0.250	0.276	0.91	0.275	0.450	0.45
5	0.200	0.218	0.92	0.22	0.372	0.37
6	0.167	0.178	0.93	0.185	0.318	0.31
7	0.143	0.152	0.94	0.16	0.278	0.275
8	0.125	0.133	0.94	0.14	0.244	0.245
9	0.111	0.117	0.95	0.12	0.218	0.22
10	0.100	0.105	0.96	0.11	0.197	0.20
15	0.066	0.068	0.97	0.075	0.133	0.14

(a) is calculated according to formula (1), (b) and (d) according to Maxwell, (c) according to formula (6) and (e) according to formula (8).

Approximation for other electrode configurations

If, continuing in this way, an attempt is made to calculate the fields for other configurations of charged electrodes, in the case of rigorous solutions the difficulties are encountered which were discussed at the beginning of the foregoing section. For example in the case of a sphere or a cylinder with a plane plate, the potential on the sphere and on the plate has a given value, and it is clear that the solutions found: $\varphi = A/r + B$ and $\varphi = A \ln r + B$ on the plate, for which r changes from point to point can never be satisfactory. By means of certain devices an adaptation of the solution to the boundary conditions is sometimes successful, and very much work has been done in this direction, although the mathematical devices rapidly become more elaborate than is desirable for practical work. For this reason the practical constructor makes little use of these results as a rule.

It is, however, possible to discuss in a simple way several important cases which give usable results on the basis of those already obtained, by means of suitable appoximations. *Such approximations are possible because in general with a given potential difference between two or more electrodes the maximum field strength occurring at the surface of one of the electrodes is not determined, or if so to only a reasonably small degree, by the shape of the other electrodes.* Because of this, with a given potential difference the maximum field strength on a cylinder coaxial with another cylinder does not differ too much for example from the field strength of the cylinder with a plane plate. This is shown in *table II* where E again represents the field strength in the case of plane plates a distance a apart and E_m the maximum field strength in the cases being compared with it.

In *fig. 2* this coaxial cylinder is drawn as a dotted line through point C of the plane plate. The plane plate was an equipotential plane, and it is therefore clear that if it is replaced by the coaxial cylinder a higher value will be found for the field strength on the inner cylinder. The surface clements

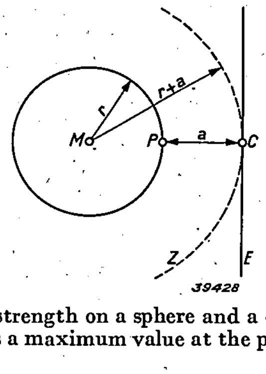


Fig. 2. The field strength on a sphere and a cylinder opposite a plane plate E has a maximum value at the point P.

Table II

Efficiency factor $\eta = E/E_m$ for fields on a cylinder surrounded coaxially by a second cylinder, compared with a cylinder with a plane plate and with parallel cylinders at a distance from each other.

$\frac{r+a}{r}$	coaxial cylinders (a)	cylinder and plane plate			parallel cylinders	
		rigorous (b)	$k = \frac{(a)}{(b)}$	$k = 0.9$ (c)	rigorous (d)	$k = 0.9$ (e)
1.5	0.811	0.861	0.94	0.90	0.924	0.96
2	0.693	0.760	0.92	0.77	0.861	0.90
3	0.549	0.623	0.88	0.61	0.760	0.77
4	0.462	0.533	0.86	0.52	0.682	0.68
5	0.402	0.468	0.86	0.45	0.623	0.61
6	0.358	0.419	0.86	0.40	0.574	0.56
7	0.324	0.380	0.85	0.36	0.533	0.52
8	0.297	0.349	0.85	0.33	0.497	0.485
9	0.275	0.323	0.85	0.31	0.468	0.45
10	0.256	0.301	0.85	0.285	0.442	0.43
15	0.193	0.228	0.85	0.215	0.349	0.33

(a) is calculated according to formula (2), (b) and (d) according to a rigorous method, (c) according to formula (3) and (e) according to formula (4).

of the outer cylinder lie closer to the inner cylinder than do those of the plane plate. Upon a comparison with the results of the rigorous calculation, however, it is found that the differences in the field strengths between the two coaxial cylinders and between a cylinder and the plate which is tangent to Z at C are relatively small. *Table II* gives the results. As was to be expected, in the case of plane plate and cylinder the field strengths are somewhat lower than for the substitution arrangement of two coaxial cylinders. The difference is, however, slight. The agreement becomes still better when a correction factor $k < 1$ is introduced into the formula. This is also shown in table II. This factor should increase slowly with increasing value of r with a constant, and for $r \gg a$ it should approach unity, since the field is then practically homogeneous. For practical purposes we choose the value 0.9. In this way one finds for the maximum field strength on the cylinder with radius r at a distance a from a flat plane

$$E_m = \frac{9}{10} \cdot \frac{U}{r \ln \dfrac{r+a}{r}} \quad . \quad . \quad . \quad . \quad (3)$$

For two parallel cylinders both with radius r at a distance a from each other, one of which is at the potential φ_1 and the other at φ_2 the maximum field strength on one of the two cylinders can be approximated by a similar line of reasoning. The plane which bisects perpendicularly the shortest distance between them is an equipotential plane with the potential $(\varphi_1 + \varphi_2)/2$. This plane lies at a distance $a/2$ from one cylinder. If one imagines this plane again replaced by a coaxial cylinder at the potential of the equipotential plane in question, one finds for the maximum field strength on the cylinder:

$$E_m = \frac{9}{10}\,\frac{\varphi_1-\varphi_2}{2}\cdot\frac{1}{r\ln\dfrac{r+a/2}{r}} = \frac{9}{20}\,\frac{U}{r\ln\dfrac{r+a/2}{r}} \cdot \quad (4)$$

Let us consider the case of two cylinders which are not parallel but perpendicular to each other. Apparently the maximum field strength, which produces breakdown, occurs at the surface of the smaller cylinder (radius r), at the point where the distance between the cylinders is shortest (P_1). It has been found experimentally that the potential difference at which breakdown occurs in the case of unlike mutually perpendicular cylinders at a distance a from each other (*fig. 3*)

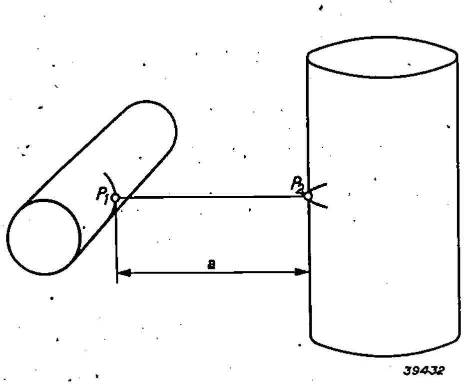


Fig. 3. Two mutually perpendicular cylinders have the greatest field strength on the cylinder with the smaller radius at point P_1 which lies closest to the second cylinder. This field strength is of approximately the same value as that for two parallel cylinders at the point of intersection of a line perpendicular to the axes with the cylinder of smaller radius.

does not differ very much from the case of parallel cylinders at a distance a from each other. Thus for crossed cylinders at a distance $P_1P_2 = a$ it is possible to write for the greatest field strength that occurs on the cylinders, as an approximation:

$$E_m = \frac{9}{20}\cdot\frac{U}{r\ln\dfrac{r+a/2}{r}} \quad . \; . \; . \; . \quad (5)$$

It is obvious from the foregoing that in the case of a sphere (radius r) at a distance a from a plane plate the following may be written for the maximum field strength on the sphere:

$$E_m = \frac{U}{a}\,\frac{r+a}{r}.$$

The accuracy of this approximation can be verified in table I, columns (a) and (b). As shown in column (c) the agreement can here also be made better by introducing a constant correction factor k, and in this case also 0.9 is seen to be good average value for the constant, so that for a sphere with a plane we may write

$$E_m = \frac{9}{10}\,\frac{U}{a}\,\frac{r+a}{r} \quad . \; . \; . \; . \; . \; . \; . \quad (6)$$

Two spheres both of radius r at a distance a from each other may again in a similar way be derived from the case of a sphere and a plane at a distance of $a/2$. Corresponding to (6) one then immediately obtains

$$E_m = \frac{9}{10}\,\frac{U}{a}\,\frac{r+a/2}{r} \quad . \; . \; . \; . \; . \quad (7)$$

For the usefulness of this approximation see table I columns (d) and (e).

The field distribution between two rings (tori) lying in the same plane may be derived from that between two parallel cylinders, like that of a ring and a cylinder whose axes are parallel.

The maximum field strength in the second case, when the cylinder and ring have the same radius, is on the ring, since the ring exhibits curvature in two directions at the point where the distance to the cylinder is smallest.

If the distance from ring to cylinder or from ring to ring is a, while the radius of both is r, we find:

$$E_m = \frac{U}{2r\ln\dfrac{r+a/2}{r}} \quad . \; . \; . \; . \quad (8)$$

No correction need be introduced here since due to the double curvature of the ring the field strength on it is greater than on the cylinder.

The case of a cylinder, for instance a wire which passes through a circular opening, can be derived from that of a ring through which a coaxial cylinder passes. It is always possible and also desirable to round off the edge of the opening so much that the maximum field strength occurs at the cylinder which thus has the smaller radius and the greater curvature. If one imagines the ring which encloses the cylinder to be replaced by a coaxial cylinder with the radius of the inside of the ring, one again obtains

$$E_m = \frac{U}{2r\ln\dfrac{r+a/2}{r}} \quad . \; . \; . \; . \quad (9)$$

Another method of approximation leads to the formula

$$E_m = \frac{U}{r \ln \dfrac{r+a}{r}}, \quad . \; . \; . \; . \quad (10)$$

which corresponds to (9) for $a \ll r$.

Influence of irregularities on the surface of the electrodes

In all the cases discussed until now the surfaces of the electrodes must be absolutely smooth. A local depression or projection leads to changes in the field strength. In a first approximation the effects of such irregularities can be seen by calculating the field strength in the neighbourhood of a hemispherical projection on a flat plane at a great distance from which there is a parallel plane. To do this we first calculate the disturbance caused by a metal sphere placed in a homogeneous field.

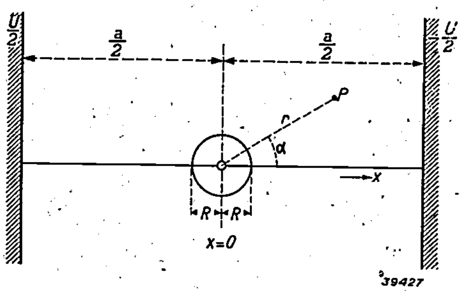


Fig. 4. A conducting sphere in a homogeneous field with the field strength E between two plane plates at potentials $-U/2$ and $+U/2$ has a maximum field strength of $3E$ on its surface.

It can be shown that upon placing a conducting sphere of radius R in a homogeneous electric field (*fig.* 4) the potential is given by

$$\varphi = -Ex + E\frac{R^3}{r^3}x. \quad . \; . \; . \; . \quad (11)$$

E is here the field strength, $-Ex$ the potential of the homogeneous field between two plates at distances $+a/2$ from the centre of the sphere. In this a is considered to be large compared with the radius R of the sphere; $\varphi = +U/2$ for $x = -a/2$, and $\varphi = -U/2$ for $x = +a/2$. The second term in formula (11) is equal to the potential of a dipole imagined to be situated at the centre [3]) which has its positive pole toward the right along the x-axis and which has a moment of R^3E. It is easily seen that for $r = R$ the potential $\varphi = 0$, *i.e.* constant, while for $r \gg R$ only the homogeneous field remains. The solution thus satisfies the boundary conditions and of course Laplace's equation.

[3]) By this is meant a charge q at the point $x = \delta$, combined with a charge $-q$ at $x = -\delta$, where $\delta \ll R$. The moment of the dipole is $2q\delta$.

The lines of force strike the metal surface perpendicularly. It is therefore sufficient to calculate the radial force E_r at the point $r = R$. If we set $x = r \cos \alpha$, then

$$E_r = -\frac{\partial\varphi}{\partial r} = E\cos\alpha + \frac{2\,ER^3}{r^3}\cos\alpha \quad . \quad (12)$$

or, for $r = R$:

$$E_r = 3E\cos\alpha.$$

The greatest field strength thus occurs where $\alpha = 0$ or 180°. It amounts to

$$E_m = 3E.$$

It is thus clear that due to the presence of the sphere in the field the maximum field strength is increased by a factor 3. The calculation also holds when only the right-hand half of fig. 4 is considered and the plane $x = 0$ is considered to be conducting. For a spherical irregularity on a charged plane plate from which lines of force emerge, the maximum field strength will therefore also become greater by a factor 3. Every irregularity, even though it is not truly spherical, gives a similar increase. Thus for a semicylinder on a plane an increase of the maximum field strength by a factor 2 has been calculated. Such an increase may be expected in the case of an elongated irregularity (scratches). We give only a diagram (*fig.* 5) of the field distribution for a spherical irregularity; that for a cylindrical irregularity shows many similarities.

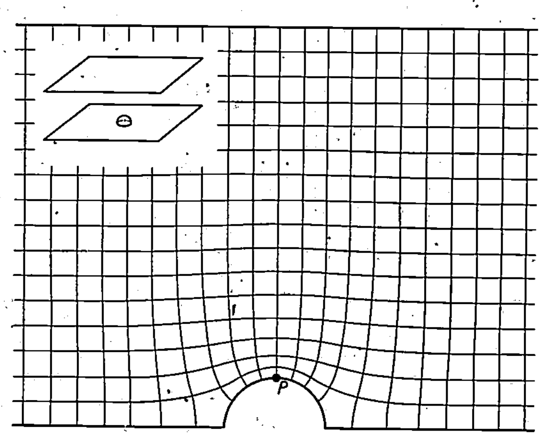


Fig. 5. Curvature of equipotential lines and lines of force in the field between two parallel plates when there is a projection in the shape of a hemisphere on one plate. At point P there is a field strength three times as large as in the homogeneous field. In the case of a half cylinder on a plate the maximum field strength at P is twice as great as in the homogeneous field.

The role of the dielectric

Until now we have not spoken of the rôle of the dielectric, since our calculations were always carried out with a dielectric constant 1. If in the problems discussed all of the space were filled with a substance with the dielectric constant ε, the field strength would not be changed for the same potential distribution. The dielectric displacement $D = \varepsilon E$ is, however, in this case everywhere increased by a factor ε. Corresponding to this increase is a change in capacity by a factor ε.

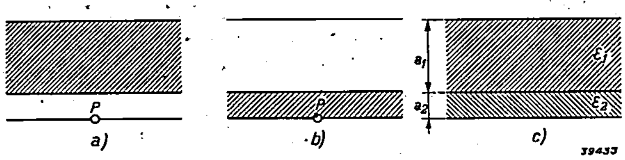


Fig. 6. Different cases of media in which the dielectric constant is not everywhere equal to unity.

If, however, space is only partly filled with a substance with a dielectric constant differing from unity, the field strength changes. Different cases hereby occur. We shall here confine ourselves to those in which the boundary surfaces are parallel plane plates (*fig. 6*) or concentric cylinders (*fig. 7*).

a) In the neighbourhood of a point P (see fig. 6) let $\varepsilon = 1$. If $\varepsilon > 1$ everywhere else, then for a dielectric bounded by plane plates the field strength at P is ε times as great, because the dielectric displacement along the lines of force is constant. A related case occurs when there are air bubbles in an insulator, for instance in the oil in a transformer. Due to the spherical form, however, the factor by which the field strength is multiplied is only $3\varepsilon/(1 + 2\varepsilon)$. Due to the greater field strength breakdown may occur in the air bubbles, accompanied by ionization, and thus depreciation of the oil.

b) In the converse case where $\varepsilon > 1$ in the neighbourhood of P and $\varepsilon = 1$ in by far the largest part of space, the dielectric displacement D changes only slightly and E_1 in the insulator becomes a factor ε smaller. The field strength E_2 in the rest of space changes only slightly.

c) In the general case of two insulators with $\varepsilon = \varepsilon_1$ and $\varepsilon = \varepsilon_2$ the capacity per cm² is determined by the two capacities C_1 and C_2 according to the formula

$$1/C = 1/C_1 + 1/C_2,$$

where $C_1 = \varepsilon_1/4\,\pi a_1$ and $C_2 = \varepsilon_2/4\,\pi a_2$.

The voltages on the layers are distributed inversely proportional to the capacities, thus

$$U_1 : U_2 = C_2 : C_1 = \varepsilon_2 a_1 : \varepsilon_1 a_2.$$

If we assume that $U = U_1 + U_2$, then

$$U_1 = U\frac{a_1\,\varepsilon_2}{a_1\,\varepsilon_2 + a_2\,\varepsilon_1} \quad \ldots \quad (13)$$

and

$$U_2 = U\frac{a_2\,\varepsilon_1}{a_1\,\varepsilon_2 + a_2\,\varepsilon_1} \quad \ldots \quad (14)$$

The field strengths in the layers are ($U/a = E$)

$$E_1 = \frac{U\varepsilon_2}{a_1\,\varepsilon_2 + a_2\,\varepsilon_1} \text{ and } E_2 = \frac{U\varepsilon_1}{a_1\,\varepsilon_2 + a_2\,\varepsilon_1} \ldots (15)$$

If $\varepsilon_2 = 1$, the field strength in the two layers becomes

$$E_1 = \frac{U}{a_1 + a_2\,\varepsilon_1} \text{ and } E_2 = \frac{U\varepsilon_1}{a_1 + a_2\,\varepsilon_1} \ldots (16)$$

The greatest field strength is E_2. For $a_1 \gg a_2$, E_2 becomes equal to $U\varepsilon_1/a_1$, for $a_1 \gg a_2$, $E_2 = U/a_2$. These are the cases discussed under *a*) and *b*).

In more complicated cases of stratified insulators between parallel plane conductors the value of E may be found by remembering that everywhere $D = \varepsilon E$ and that the voltages are distributed over the different layers according to the capacities. The determination of the field strength then always takes place in the way indicated for case *c*).

Several interesting configurations deserve closer consideration. In fig. 7 the case of concentric cylinders is represented where the dielectric is divided into layers. We have seen that the field strength increases as $1/r$ from the outside inwards. If the values $\varepsilon_1, \varepsilon_2, \varepsilon_3 \ldots \varepsilon_n$ are chosen so that $\varepsilon_1 r_1 = \varepsilon_2 r_2 = \ldots \varepsilon_n r_n$, the field strength in the layers is about equal. The condensers formed by the different layers then have the same capacities per unit of surface:

$$\frac{2\pi r\varepsilon}{4\pi d} = \frac{r\varepsilon}{2d},$$

where d is the thickness of layer. The same voltages thus act. Such an arrangement of dielectrics is often proposed for cables.

Summarizing it may be stated that by the introduction of insulators into an electric field the maxi-

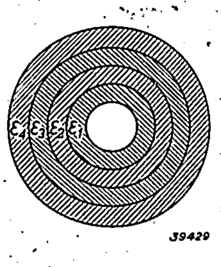
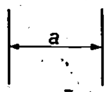
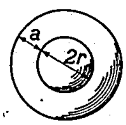
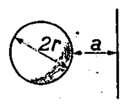
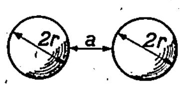
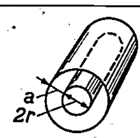
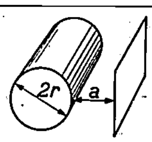
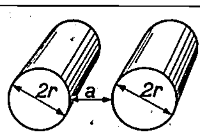
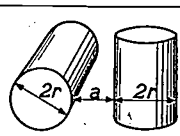
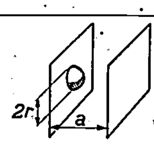
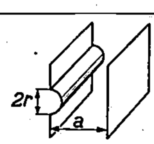
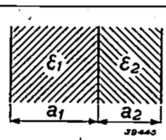


Fig. 7. Different dielectrics in coaxial layers.

Table III

Maximum field strength E with a potential difference U between the electrodes, for different electrode configurations.

Configuration		Formula for E	Example
Two parallel plane plates		$\frac{U}{a}$	$U = 100$ kV, $a = 2$ cm, $E = 50$ kV/cm.
Two concentric spheres		$\frac{U}{a} \cdot \frac{r+a}{r}$	$U = 150$ kV, $r = 3$ cm, $a = 2$ cm, $E = 125$ kV/cm.
Sphere and plane plate		$0.9\,\frac{u}{a} \cdot \frac{r+a}{r}$	$U = 200$ kV, $r = 5$ cm, $a = 8$ cm, $E = 58.5$ kV/cm.
Two spheres at a distance a from each other		$0.9\,\frac{U}{a} \cdot \frac{r+a/2}{r}$	$U = 200$ kV, $r = 5$ cm, $a = 12$ cm, $E = 33$ kV/cm.
Two coaxial cylinders		$\frac{U}{2.3\; r\; \lg \frac{r+a}{r}}$	$U = 100$ kV, $r = 5$ cm, $a = 7$ cm, $E = 22.9$ kV/cm.
Cylinder parallel to plane plate		$0.9\,\frac{U}{2.3\; r\; \lg \frac{r+a}{r}}$	$U = 200$ kV, $r = 5$ cm, $a = 10$ cm, $E = 32.8$ kV/cm.
Two parallel cylinders		$0.9\,\frac{U/2}{2.3\; r\; \lg \frac{r+a/2}{r}}$	$U = 150$ kV, $r = 6$ cm, $a = 20$ cm, $E = 11.5$ kV/cm.
Two perpendicular cylinders		$0.9\,\frac{U/2}{2.3\; r\; \lg \frac{r+a/2}{r}}$	$U = 200$ kV, $r = 10$ cm, $a = 10$ cm, $E = 22.2$ kV/cm.
Hemisphere on one of two parallel plane plates		$\frac{3U}{a}$; $(a \gg r)$	$U = 100$ kV, $a = 10$ cm, $E = 30$ kV/cm.
Semicylinder on one of two parallel plane plates		$\frac{2U}{a}$; $(a \gg r)$	$U = 200$ kV, $a = 12$ cm; $E = 33.3$ kV/cm.
Two dielectrics between plane plates $(\varepsilon_1 > \varepsilon_2)$		$\frac{U\,\varepsilon_1}{a_1\,\varepsilon_2 + a_2\,\varepsilon_1}$	$U=200$ kV, $\varepsilon_1=2$, $\varepsilon_2=4$, $a_1=6$ cm, $a_2=5$ cm, $E = 11.8$ kV/cm.

mum field strength which existed in air is considerably increased, and in extreme cases (thin layer of air) proportionally to ε, but in general less.

Examples

In *table III* a number of cases are collected of configurations of electrodes which occur commonly in the construction of apparatus for high voltages.

In addition to the formula valid for the configuration an example is given for all cases. All the formulae are derived in Briggs logarithms (to the base ten) since many persons are more accustomed to their use than to the use of tables of Napierian logarithms. We had this chart in mind when we spoke of the prescriptions which may be used to guide the constructor. Given a definite structural design, it is possible with the help of this table to calculate for every point the maximum electrical field strengths occurring from the known dimensions and potentials of the structural components. A comparison of this field strength with the breakdown voltages of the surrounding dielectrics shows the constructor what are the dangerous spots in his design. Devices for the improvement of such spots are, in addition to changes in shape (consisting for example in the increase of the radii of curvature of the components), a possible replacement of the dielectric by a different one according to the rules given above. Sometimes, as already mentioned, the electrode can be surrounded by one with a larger radius of curvature.

Although the chart given furnishes an idea of the maximum field strength occurring for a large number of shapes of electrodes, it is by no means complete. Even the case of two parallel plane plates, where $E = U/a$ can, moreover, only be correct when the plates extend to infinity. If the edges are not well rounded the first breakdown occurs at that spot. This can be avoided by choosing curved instead of plane surfaces, so that their distance apart becomes greater toward the edges [4]).

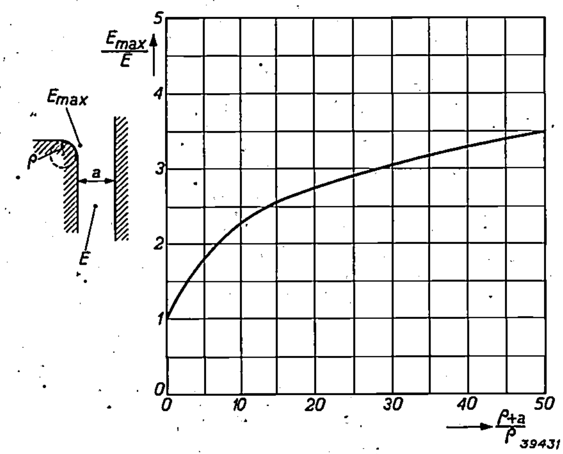


Fig. 8. Plane plate with bent edge opposite an infinite plane plate. Influence of the curvature of the bent edge on the maximum field strength.

A problem which will consequently sometimes occur in practical cases is the field distribution of a round corner opposite a plane plate. The solution of this is also by no means simple [5]). Because of the importance of this case we give the solution of this problem in a graph (*fig. 8*), which indicates values of E_{max}/E for different values of $(\varrho+a)/\varrho$ ($a =$ distance between the plates, $\varrho =$ radius of curvature of the rounded edge). $E = U/a$ is here the field strength between the plates, and it may be seen from the graph that for small values of ϱ, E_{max} may be more than 3.5 times as large as E.

[4]) W. Rogowski and H. Rengier, Arch. Elektrotechn. **16**, 73, 1926.

[5]) A. Dreyfus, Arch. Elektrotechn. **13**, 123, 1924, see fig. 10.

EXPERIMENTS ON THE AUSTEMPERING OF STEEL

by J. G. C. STEGWEE.

621.785.6

In the process of hardening steel the cooling of the heated object must take place so rapidly that the austenite (unstable below 720° C) is not converted into the soft perlite, which is formed at temperatures above ± 500° C, but into the hard martensite formed below 150° C. A closer investigation of the transformation has shown, however, that when it takes place in the temperature range from 200 to 400° C a structure with remarkable mechanical properties is formed (bainite): its hardness is indeed less than that of martensite, but its toughness is considerably greater. The process used to obtain this transformation product is called austempering. Several experiments are described which where carried out with this method of hardening on a special spring steel and which produced good results. For other applications also, for instance in the caps of the "Philishave" electric shaving apparatus, for which the material, in addition to a high resistance against wear, must possess a certain toughness, austempering produces better results than the ordinary method of hardening.

The hardening of steel

In the ordinary process of hardening steel which has been used for centuries, the object to be tempered is first heated to a high temperature, for instance 800° C, whereupon the iron is completely transformed into γ-iron in which the carbon present is dissolved. In this way the so-called austenite structure is formed. Below a temperature of 720° C γ-iron is not stable, upon cooling therefore transformation takes place into α-iron in which carbon is only slightly soluble. The manner of separation of the carbon and thus the structure obtained depends, not only on the content of carbon and other components, but also very much upon the rate of cooling. With slow cooling a more or less coarse lamellar structure (perlite) is in general obtained, which is soft; upon rapid cooling, on the other hand, an irregular structure of fine crystalline needles occurs (martensite), which gives the material the desired great hardness. *Fig. 1* shows photographs of etched surfaces of the various structures mentioned.

A fairly recent investigation by Davenport and Bain [1]) has made it possible to define more precisely the concepts "slow" and "rapid" cooling. These authors immersed test rods of unalloyed steel with 0.78 per cent of carbon [2]), after heating above 720° C, in different baths of molten salts or alloys whose temperatures lay between 0 and 720° C. Each rod was thus quenched to a given temperature below the transformation point and then kept at that temperature. The transformation of the austenite is then found to proceed imperceptibly slowly at first, probably because a sufficient number of nuclei of the new crystal structure must first be formed, and only after a definite time t does the conversion proceed to an appreciable extent. After a time t_2 the conversion is completed. The important point in this process is, that not

[1]) Trans. Amer. Inst. Mining Metallurg. Eng. (Iron & Steel), 1930, p. 117.

[2]) This carbon content is chosen because at this content there is aneutectic point of the iron-carbon system; at other contents of carbon, upon cooling the steel, a separation of iron or carbon already takes place above the transformation point of the austenite.

39320

a b c

Fig. 1. Etched surfaces of austenite (a), perlite (b) and martensite (c). The photograph (a) is of a special kind of steel whose austenite structure decomposes only after a long time even at room temperature. The successive laminae in the perlite structure (b) consist alternately of ferrite (pure α-iron) and cementite (Fe_3C).

only the structure finally formed (and thus the hardness), but also the times t_1 and t_2 depend very closely upon the temperature T of the salt bath, *i.e.* on the temperature at which the transformation takes place.

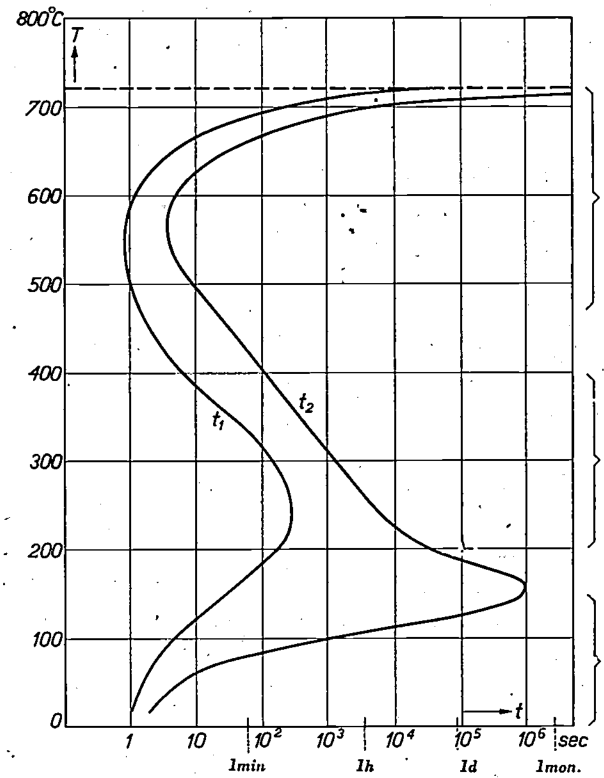


Fig. 2. S-shaped region in which the transformation takes place of the austenite unstable below 720° C. The curves were recorded by the laboratory of the U.S. Steel Corp. (Metal Progress, 1939, p. 374) for an unalloyed steel with 0.78% of carbon. For each temperature T the time t_1 of the beginning and t_2 of the end of the transformation is plotted. On the right the transformation products obtained at the different temperatures are indicated: at high temperatures perlitic structures are formed, below 150° C martensite.

In *fig. 2* the results of this investigation are summarized. Although from the nature of the case the limits and t_2 are not sharp, an S-shaped transformation region is clearly indicated. At temperatures only slightly lower than the transition point, for instance 700° C, the transformation begins only several minutes after quenching, and is complete after about one hour. At temperatures between 500 and 600° C the transformation begins almost immediately and is finished in several seconds. At still lower temperatures, for example from 200 to 300° C, the beginning of the transformation is again very much retarded (5 to 10 minutes) and the transformation occupies several hours. Finally at temperatures in the neighbourhood of room temperature the beginning and end of the transformation again occur very quickly after quenching (several seconds to 1 minute).

In fig. 2 the nature of the transformation product obtained is also indicated. At temperatures only slightly below the transition point, coarse lamellar perlite is obtained; as the transformation temperature becomes lower the perlite becomes finer in structure, until at 500° C the lamellar structure is hardly recognizable any longer. Below about 150° C the austenite is transformed into martensite.

If after having been made austenite by heating, the steel is s l o w l y cooled (normalized), the temperature of the object will vary for example according to line *1* in *fig. 3*. The transformation of the austenite will begin at the temperature T_a and end at the temperature T_b so that the result will be a structure which is a mixture of the soft perlitic structures which correspond according to fig. 2 to the temperature range between T_a and T_b. Upon cooling the heated steel very quickly on the other hand, the temperature of the object varies for instance according to line *2*, the transformation of the austenite now takes place at the low temperatures between T_c and T_d, and a h a r d structure of martensite is obtained. Therefore in order to harden the steel it must obviously be cooled so quickly that the temperature region between 500 and 600° C is traversed before the transformation can start.

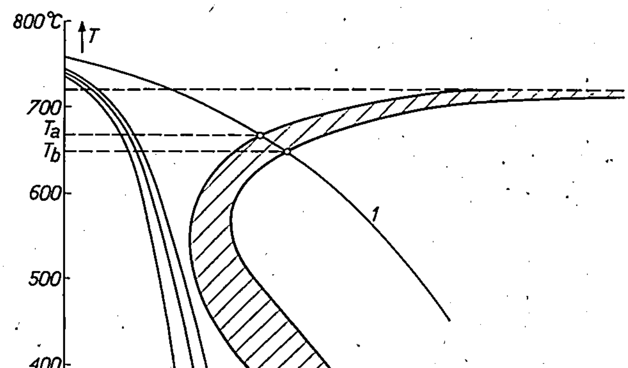


Fig. 3. S-shaped transformation region for a certain kind of steel and schematic variation of temperature upon the use of different methods of cooling the object. *1*, slow cooling (normalizing); the transformation of the austenite takes place at temperatures between T_a and T_b, perlite is formed. *2*, rapid cooling (ordinary tempering); transformation at temperatures between T_c and T_d, martensite is formed. *3*, stepped hardening. *4*, austempering.

The position of the S-shaped transformation region is not the same for all kinds of steel, and the rates of cooling required for hardening differ accordingly. There are kinds of steel for which the dangerous curve of the S is shifted so far to the right that ordinary cooling in air is already "rapid" enough to produce a hard product; in the case of other kinds of steel quenching in oil instead of water is sufficient (due to the higher boiling point and the poorer heat conduction and convection of oil, cooling in oil is much less sudden). In the case of large objects this possibility of a more gradual cooling is very welcome. Upon quenching in water it is only the outer layers of the material which can immediately be cooled to temperatures below 100° C, while the cooling of the inner parts always experiences a certain retardation. This may result in the fact that the inner material does not become hard enough, and due to the uneven cooling and hardening all kinds of stresses and deformations may also occur in the object, expecially since the transformation from austenite into martensite is accompanied by a volume change (expansion of about 1 per cent). Ordinarily the stresses occurring in steel hardened in water are reduced by heating the object for some time after hardening to a temperature of from 150° to 300° C (tempering), whereupon, however, the hardness is somewhat diminished.

Another possibility of combatting strains is the so-called stepped hardening, which has been employed for several decades. In this process the hot object is quenched in a salt bath of for instance 200° C, left in it for several minutes and then slowly cooled further in air. From fig. 3 in which stepped hardening is represented by line *3* it may clearly be seen that also in this case a martensite structure is obtained here when the object is not left too long in the salt bath. Since during this interval before the transformation the object assumes a somewhat more uniform temperature, and then cools further only gradually, the stresses are a great deal smaller. In this case thus the annealing precedes as it were the actual hardening.

Austempering

The American investigations mentioned above have not only given a deeper insight into the different hardening processes, but have also led to the development of a new method of hardening, "austempering". Upon a closer consideration of the different structures which are obtained upon the transformation of austenite at different temperatures, it was found that the transformation product obtained between about 200 and 400° C deserved special attention. Until that time it has been taken simply for very fine lamellar perlite; compared with perlite, however, it was found to possess much greater hardness, namely 400—700 Vickers compared with 200—300 for perlite, while at the same time the toughness of the material was found to be considerably greater than martensite which is tempered to the same hardness. These properties justified the expectation that in some cases where the relatively low hardness is sufficient the martensite could be better replaced by the transformation product in question, which has been given the name "bainite".

Indeed the bainite structure has been put into practical use in America during the last few years. The method of obtaining it is immediately clear after the foregoing: the hot steel is quenched in a salt bath of a temperature of, say 300° C, as is done in stepped hardening. But in this case the steel is left in the bath until the austenite is completely transformed, see line *4* fig. 3. The further cooling may take place rapidly or slowly, the properties of the material show no further appreciable change.

In the Philips factories several experiments have been carried out recently to investigate the practical applicability of this method of tempering with different kinds of steel. In the first place a steel was investigated which is used for making springs, and whose composition is given in *table I*. For this use the toughness of the material is especially important.

Table I

Composition of spring steel for which austempering was tried.

carbon	0.6—0.7%
silicon	0.35%
phosphorus *)	<0.03%
sulphur *)	<0.03%
manganese	0.60%
iron	remainder

*) phosphorus and sulphur together <0.05%.

In order to apply the method of austempering to this material it was first necessary to determine the times necessary for the transformation of austenite at different temperatures, *i.e.* the boundaries of part of the S-shaped transformation region lying between 200 and 400° C had to be found as was done by Davenport and Bain [1]). For this purpose a number of strips of $70 \times 8 \times 0.8$ mm of the steel to be investigated were heated (this was done in an atmosphere of nitrogen in order to limit decarburization and oxidation), and then all were immersed together in the salt bath of the temperature T. After different times t a single strip was taken out of the bath and quenched in water. By means of a hardness test the degree could then be determined to which the austenite had been transformed. As long as the times in the salt bath are so short that the conversion of the austenite has not yet begun and the strip in question therefore still consists entirely of austenite, the latter is entirely transformed into martensite upon being quenched in water, and thus attains practically the greatest possible hardness (about 840 Vickers). If a lower hardness is measured it means that in addition to the martensite there is also a softer component present (in this case bainite), and that therefore the transformation of the austenite had already begun in the salt bath. The lower the hardness the farther the conversion had proceeded, until, when the hardness no longer decreases, it may be assumed that the transformation was complete and that the product obtained contains no martensite but only bainite. In *fig. 4* the results of these experiments for a series of temperatures T of the salt bath are represented graphically by plotting the measured hardness as a function of the time t of remaining in the salt bath. From each curve the times t_1 of the beginning and t_2 of the end of the conversion can easily be read off.

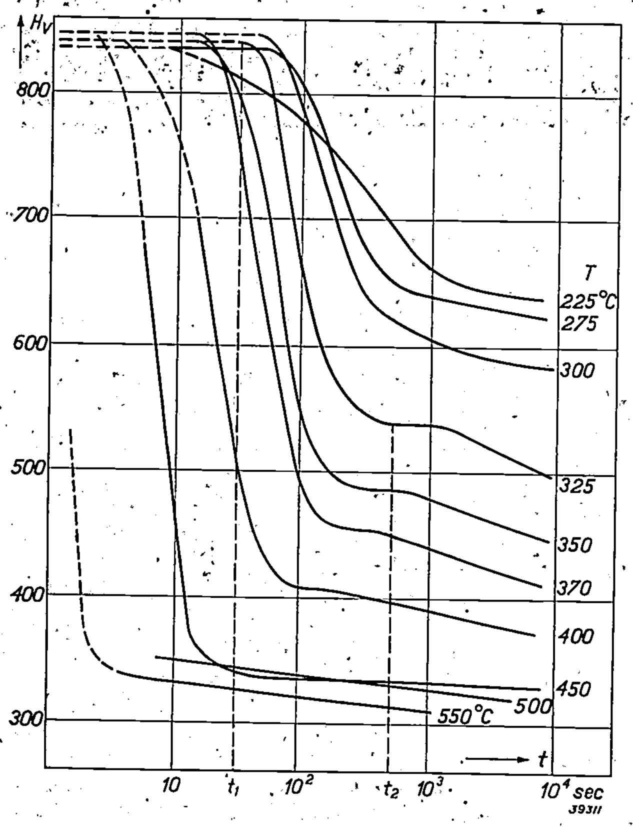


Fig. 4. Hardness measurements of strips of spring steel for determining the S-shaped transformation region. The strips which were quenched in a bath of the temperature T, where each, after a certain time i quenched further in water, and the hardness H_V in Vickers units was then measured. H_V is plotted as a function of t with the temperature T as a parameter. From each curve the time t_1 of the beginning and the time t_2 of the conclusion of the transformation can be read off.

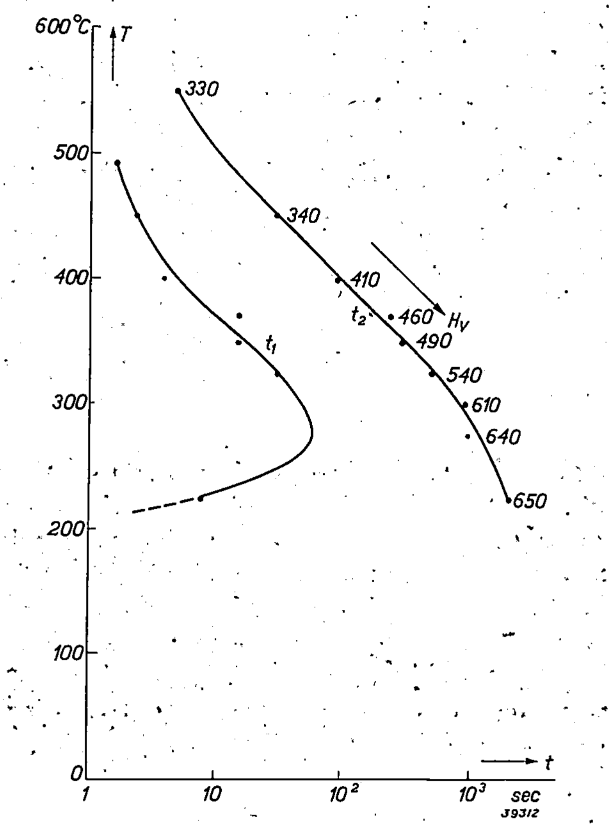


Fig. 5. Transformation region of spring steel. The times t_1 and t_2 from fig. 4 are plotted against the temperature T, while the final hardness obtained H_V (upon complete transformation into bainite) is also indicated.

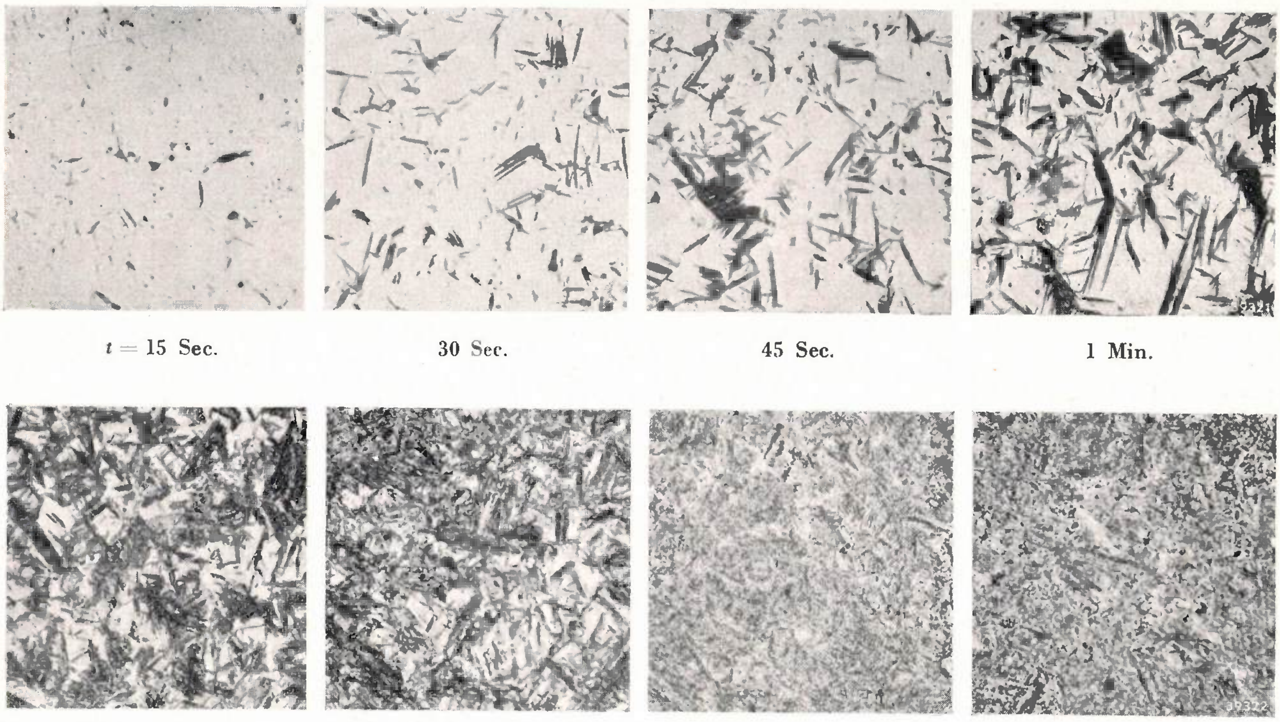
In *fig. 5* the times t_1 and t_2 deduced from fig. 4 are plotted as functions of the temperature T. In this way the part of the S-region for the spring steel in question, which is of importance in austempering is determined. At the same time the hardness which was finally obtained at each temperature is indicated in the figure.

The course of the transformation may also be followed by microscopic examination of the structure of the strips quenched in water, instead of by hardness masurements. *Fig. 6* shows photographs of the structure of eight strips taken out of the salt bath after different times, at a temperature of the salt bath of 350° C. The light-coloured component in these photomicrographs is martensite, the darker component is the more rapidly etched bainite.

In order to ascertain how the austempered spring steel will behave in practice the toughness and the resiliency was determined of a number of strips which were austempered at different temperatures and thus possessed different hardnesses. The resiliency, which was determined with Tarnogrocki's apparatus by bending the test strip 180° around a mandrel of 20 mm, and measuring the angle through which the strip springs back on removal of the bending force, directly expresses the "springy" qualities of the material. It is closely correlated with the height of the limit of elasticity and thus with the hardness. The toughness, which was determined by measuring the angle through which a test piece can be bent around a mandrel of 1 mm without breaking, provides a measure for the work of deformation which the spring can take up. In *table II* the results of the measurements are summarized. At the same time for the sake of comparison in *table III* the hardness, resilience and toughness are given for a series of strips of the same spring steel which were hardened in oil in the ordinary way and then tempered for 15 minutes at different temperatures.

It is found from a consideration of the two tables that the austempered material has the greater toughness throughout with a given resiliency (or hardness). The conclusions of the American investigators about the favourable effect of austempering are thus confirmed for the steel here examined. In *fig. 7*, in which the toughness is plotted against the resiliency, the superiority of austempered steel is very obvious.

Not only for springs, but also for other applications, the combination of qualities of austempered material may be useful, for example in cases where at one and the same time a given minimum toughness and the highest possible resistance to wear are required [3]). An example of this is found in the caps



t = 15 Sec. 30 Sec. 45 Sec. 1 Min.

$t = 1^1/_2$ Min. 2 Min. 3 Min. 4 Min.

Fig. 6. Photographs of etched surfaces of strips of spring steel which were quenched in a bath with a temperature of 350° C and remained in it for different times t, before being further cooled in water. The light-coloured component is martensite, the darker is the more rapidly etched bainite.

Table II

Mechanical properties of spring steel which has been austempered at different temperatures T.

T in °C	t in min.	H_V	Resiliency in °	Angle of bending in °
225	120	657	88	47
250	90	647	90	44
275	60	632	86	50
300	40	572	76	55
325	20	505	70	81
350	10	473	65	89
375	$7^1/_2$	428	59	105
400	5	399	54	about 160

[3]) It must be kept in mind that it can only be the case where very great hardnesses are not required. Tool steel, for example, which must also be tough, can certainly not be austempered, since hardnesses of more than 800 Vickers are required.

Table III

Mechanical properties of spring steel normally hardined in oil, after tempering at different temperatures.

Tempering temperature in °C	H_V	Resiliency in °	Angle of bending in °
275	639	89	27
300	599	85	86
325	556	82	39
350	546	78	43
375	512	75	49
400	485	71	52
425	450	68	54
450	428	64	58
475	401	60	67
500	366	56	73

of the "Philishave" electric shaving apparatus [4]). These caps, perforated plates less than 0.1 mm thick, must be extremely resistant to wear since the cutters pass over them as they rotate to cut off the hairs. At the same time, however, the material of the cap may not be too brittle, since otherwise upon falling or being struck there would be danger of its breaking. Experiments actually showed that in the case of the steel of which these caps are made, the simultaneous requirements of toughness and hardness could be better satisfied by austempering.

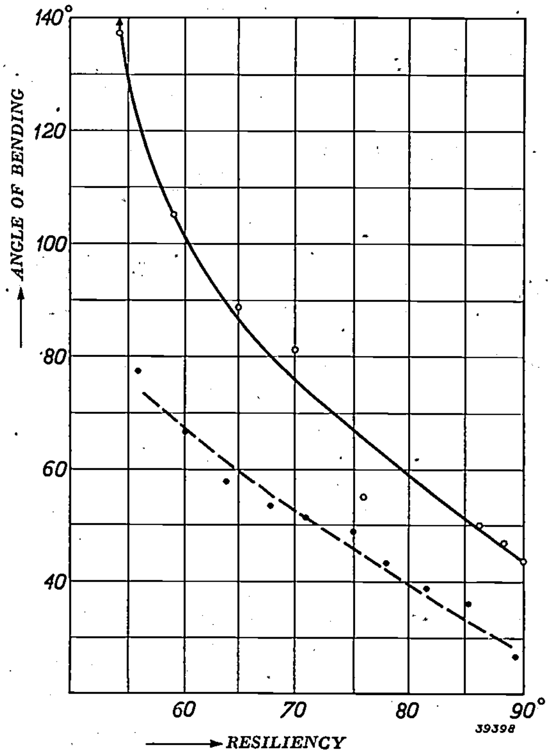


Fig. 7. For a number of normally hardened and tempered strips (brokenline curve) respectively austempering strips (drawn line) the measured toughness (angle measured in bending test) is plotted against the resiliency (angle measured with Tarnogrocki's apparatus).

4) Philips techn. Rev. 4, 350, 1939.

RECTIFICATION

In the article of J. F. Schouten, Non-linear Distortion of Sound Film with Oblique Light Slit, Philips techn. Rev. 6, 110, 1941, it is has been presumed that untill the appearance of the cited article an exact mathematical solution of this problem was still outstanding. Dr. A. Narath kindly directs our attention to the fact that already in 1936, be it on an entirely different argumentation, he published an exact solution of this problem in Kinotechnik, 18, 177—180, 1936.

VOL. 6 No. 10, p. 285-316 OCTOBER 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

THE DIODE AS A FREQUENCY CHANGING VALVE ESPECIALLY WITH DECIMETRE WAVES

by M. J. O. STRUTT and A. van der ZIEL. 621.385.2

The frequency-changing valves used in radio receiving sets for broadcasting reception are not suitable for frequency transformation on decimetre waves. At these wave lengths a diode can be used to advantage for that purpose. The conversion amplification which can be attained with a diode is always less than unity and under the most favourable conditions may approach unity. For this a high impedance of the intermediate frequency circuits is necessary, as well as a high voltage of the auxiliary oscillator. Under these conditions the input and output resistance of the connections are also large, so that the oscillation circuits in connection are little damped by the frequency-changing circuit. Upon frequency transformation by means of a higher harmonic of the oscillator voltage the conversion amplification obtained is almost as large as upon the use of the fundamental frequency. It is, however, advisable not to go farther than the fourth or fifth harmonic. In the case of the diode frequency changers, developed by us, two diodes are connected in push-pull arrangement. Furthermore the valve contains a triode which serves for the excitation of the voltage of the auxiliary oscillator. This voltage is fed to the two diodes in the same phase, whereby a balanced intermediate-frequency voltage is obtained. In order to prevent as far as possible mutual effects of the oscillation circuits in the connection, the greatest care should be devoted to the symmetrical construction of the connections. In conclusion several particulars of the construction, which have been applied in order to improve the short wave properties of these valves, are discussed.

The frequency transformation from high-frequency oscillations to intermediate-frequency oscillations customary in modern receiving sets practically always takes place in sets designed for broadcasting reception by means of a hexode, heptode or octode. In these types of valves the high-frequency oscillations received and those of the auxiliary oscillator are fed to separate grids, by which means any mutual influence of the oscillation circuits connected to these grids is avoided as far as possible. By a suitable choice of working conditions, with the frequency changers mentioned a large value of the ratio of the intermediate-frequency A.C. voltage obtained to the high-frequency A.C. voltage fed to the frequency changer (conversion amplification) can be obtained. For the reception of very short waves (decimetre waves), however, the "normal" frequency-changing valves are unsuitable. The main objection to the use of a multigrid valve as frequency changer on decimetre waves lies in the very strong damping effect which this valve exerts on the input and output oscillation circuits in connection. As has already been explained in previous articles in this periodical, the main causes of this damping are the self-induction of the connection lines to the electrodes and the transit time of the electrons [1]). The transit time of the electrons, moreover, sets an upper limit to the frequency region in which conversion amplification can be attained, just as is the case in high-frequency amplification. The aim of this article is to show that on decimetre waves a diode can be used advantageously as frequency changer.

In conjunction with this we shall give a description of a type of frequency changer for short waves developed by us in which the frequency transformation takes place by means of a diode, while at the same time a triode is also built in which serves for the excitation of the auxiliary oscillator voltage.

The diode as frequency changer

The principle of the diode frequency-changing

[1]) Philips techn. Rev. 1, 171, 1936, 3, 103, 1938.

stage is represented in *fig. 1*. The arrangement consists of a connection in series of the diode with two sources of A.C. voltage, one of which furnishes the

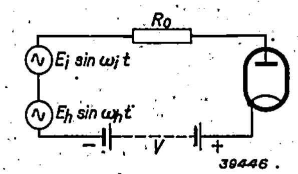


Fig. 1. Diagram showing the principle of a diode frequency changer.

high-frequency voltage, $E_i \sin\omega_i t$ and the other the voltage of the auxiliary oscillator $E_h \sin\omega_h t$. In practice E_i is always small compared with E_h. Furthermore in the series connection are included a source of D.C. voltage V and an impedance consisting of an oscillator circuit which is tuned to the difference frequency $\omega_0 = \omega_h - \omega_i$. This means that this impedance is very small for all frequencies except ω_0, for which it forms a pure resistance R_0. On the resistance R_0, therefore, only an A.C. voltage with the frequency ω_0 can act; we call this voltage the intermediate-frequency voltage E_0. The ratio of E_0 to E_i is called the conversion amplification. We shall first examine the magnitude of this conversion amplification.

In *fig. 2* the characteristic of an ordinary diode is represented. The current which flows in the diode when a D.C. voltage is applied between anode and cathode is plotted vertically. The voltage in question is indicated along the horizontal axis. Thus when a negative D.C. voltage V is applied to the diode no anode current flows in the valve. We shall now first assume that besides this D.C. voltage V only the oscillation voltage E_h acts on the diode, and that this voltage is so high that a certain anode current flows only during a very small part of a period.

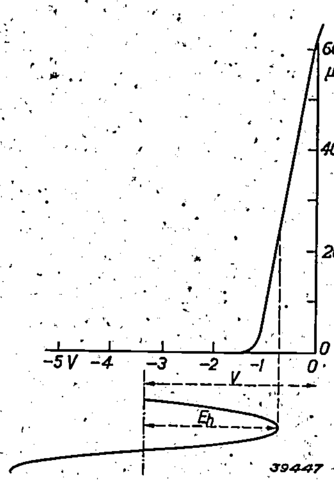


Fig. 2. Diode characteristic in which is indicated the D.C. voltage V and the oscillator voltage $E_h \sin\omega_h t$.

During the rest of each period, therefore, the anode current is zero. The variation of the voltage on the diode as a function of the time is represented in the lower part of fig. 2.

At each point on the characteristic the diode has a certain slope for a small A.C. voltage, such for example as the signal voltage E_i. From fig. 2 therefore one finds a definite variation of this slope as a function of the time. In *fig. 3* the result of such a

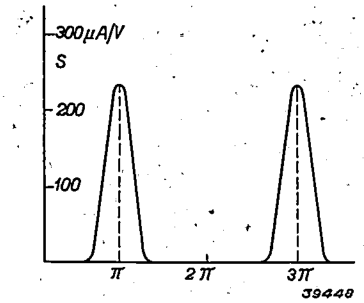


Fig. 3. Variation of the slope as a function of the time upon the application of an oscillator voltage E_h and a D.C. voltage V to a diode with a characteristic as in fig. 2.

construction is given. The slope is here plotted vertically in microamperes per volt, while the horizontal axis is a time axis on which 2π represents one period of the oscillator voltage. This variation of the slope as a function of the time can be approximated very well by a set of triangles, as shown in *fig. 4* [2]). The base of these triangles, $2b$, is smaller

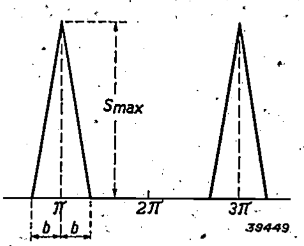


Fig. 4. Approximation of fig. 3 by triangles.

the shorter the time during which current flows in the diode, compared with a period of the oscillator voltage. Since the use of the approximation in question simplifies the following considerations exceedingly, we shall in the following assume that the variation of the slope as a function of the time takes place according to fig. 4, and we shall call the function $s(t)$. Like every periodic function $s(t)$ can be written in the form of a Fourier series which here has the form

$$s(t) = A_0 - A_1 \cos\omega_h t + A_2 \cos 2\omega_h t - A_3 \cos 3\omega_h t + \ldots \quad (1)$$

When we call the maximum slope attained in every

[2]) See M. J. O. Strutt, Diode Frequency Changers, Wireless Eng. **13**, 73, 1936.

period $s_{\max}$, we can calculate the following values for the quantities A of the series (1):

$$\left.\begin{aligned} A_0 &= \frac{1}{\pi}\, s_{\max}\, \frac{b}{2}, \\ A_1 &= \frac{2}{\pi}\, s_{\max}\, \frac{1-\cos b}{b}, \\ A_2 &= \frac{2}{\pi}\, s_{\max}\, \frac{1-\cos 2b}{4b}, \\ A_3 &= \frac{2}{\pi}\, s_{\max}\, \frac{1-\cos 3b}{9b}, \\ &\text{etc.} \end{aligned}\right\} \quad \ldots \quad (2)$$

If now in addition to the oscillator voltage E_h the high-frequency voltage E_i is also applied, an alternating current will flow in the connection which contains components with different frequencies. One of these components will have the frequency ω_0 and this will cause the occurrence on the tuned impedance R_0 of the required intermediate-frequency E_0. If the amplitude of the A.C. component with the frequency ω_0 is equal to i_0, then

$$E_0 = -i_0 R_0 \quad \ldots\ldots\ldots \quad (3)$$

The magnitude of i_0 can be found by considering that now, in addition to the oscillator voltage, two small A.C. voltages act on the diode, namely E_i with the frequency ω_i and E_0 with the frequency ω_0. The alternating current which occurs as a result of these two voltages is given by the product of the slope and the sum of the voltages mentioned, thus

$$i = s(t) \cdot (E_i \sin \omega_i t + E_0 \sin \omega_0 t). \quad \ldots \quad (4)$$

From the series for $s(t)$ only those terms are important which furnish an A.C. component with the frequency ω_0. These are, firstly, the term A_0, which multiplied by $E_0 \sin \omega_0 t$, gives $A_0 E_0 \sin \omega_0 t$, and secondly, the term $-A_1 \cos \omega_h t$, which, multiplied by $E_i \sin \omega_i t$ gives:

$$\begin{aligned} &-A_1 \cos \omega_h t \cdot E_i \sin\omega_i t = \\ &= {}^1\!/_2 A_1 E_i \,\}\sin(\omega_h - \omega_i)t - \sin(\omega_h + \omega_i)t\{. \end{aligned}$$

Since $\omega_h - \omega_i = \omega_0$, a term with the frequency ω_0 is also formed from this. The total A.C. component with this frequency is therefore:

$$i_0 \sin \omega_0 t = A_0 E_0 \sin \omega_0 t + {}^1\!/_2\, A_1 E_i \sin \omega_0 t. \quad \ldots \quad (5)$$

From (3) and (5) it follows that

$$\frac{E_0}{E_i} = -\frac{{}^1\!/_2\ A_1}{\dfrac{1}{R_0} + A_0} \quad \ldots\ldots \quad (6)$$

Since the phase of the intermediate-frequency voltage obtained is of no importance, we give as the conversion amplification a_c, the absolute value of the ratio of E_0 to E_i, thus:

$$a_c = \left|\frac{E_0}{E_i}\right| = \frac{{}^1\!/_2\, A_1}{\dfrac{1}{R_0} + A_0} \quad \ldots\ldots \quad (7)$$

With the help of this last equation we shall now make an estimate of the magnitude which a_c can attain.

According to (1) A_0 is the average value of the slope of the diode calculated over one or more periods of the oscillator voltage. It is therefore the average conductivity of the diode for every small A.C. voltage which acts in the diode in addition to the oscillator voltage, thus for example for E_i or E_0. The reciprocal value of this conductivity is called the internal resistance r_d of the diode. This quantity thus depends upon the D.C. voltage V and the oscillator voltage E_h. According to the above

$$A_0 = \frac{1}{r_d} \quad \ldots\ldots\ldots \quad (8)$$

From (2) and (8) it follows that

$$A_1 = \frac{1}{r_d}\, \frac{4(1-\cos b)}{b^2}, \quad \ldots\ldots \quad (9)$$

while by substitution of the last two equations in (7) the following is obtained:

$$a_c = \frac{R_0}{R_0 + r_d} \cdot \frac{2\,(1-\cos b)}{b^2} \quad \ldots \quad (10)$$

When, as is usually the case in practice, we are concerned with a fairly large oscillator voltage amplitude E_h (several volts) and we choose the D.C. voltage V so large that current flows in the diode only during a very small part of a period of the oscillator voltage, b is small compared to π and the second fraction in (10) approaches unity. The conversion amplification is therefore then

$$a_c = \frac{R_0}{R_0 + r_d} \quad \ldots\ldots \quad (11)$$

From this we see that a_c is always less than unity. If, however, a very large value of R_0 is used (intermediate-frequency circuit of very good quality) a conversion amplification can be obtained which differs only slightly from unity [3]).

---

[3]) Although for $a_c<1$ one may not actually speak of a conversion amplification, for the sake of uniformity we shall retain the expression for all values of a_c.

When the tuned impedance R_0 on which the intermediate-frequency voltage E_0 occurs is not included directly in the series connections as is represented in fig. 1, but *via* a transformer, a conversion amplification greater than unity can be obtained. The extent to which this increase in the conversion amplification is possible depends upon the ratio between the internal resistance, which the frequency-changing connection exhibits for intermediate frequency, and the value of R_0. Since R_0, the resonance resistance of an intermediate-frequency circuit, will often be considerably larger than the internal resistance of the frequency-changing connections for intermediate frequency, this means of obtaining a greater conversion amplification may often be applied. We shall return in the following to the calculation of the above-mentioned internal resistance.

In the reception of very high frequencies (decimetre waves) the excitation of the oscillator voltage E_h with a frequency practically equal to that of the input signal may offer difficulties, either in that the desired magnitude of several volts is not attained or that the frequency is not sufficiently stable. In such a case use is often made of an oscillator voltage with a frequency several times lower. The intermediate frequency then occurs as the difference in frequency between the input voltage and a harmonic of the oscillator voltage. In the types of the frequency-changing valves used in broadcasting receivers this method of frequency transformation under normal working conditions is always accompanied by a considerable loss of conversion amplification. It is, however, easy to understand that this need not be the case upon the use of a diode. If we apply this method of frequency transformation, then $\omega_0 = 2\omega_h - \omega_i$ or $\omega_0 = 3\omega_h - \omega_i$, etc. If in equation (4) we again take only those terms of the series for $(s(t)$ which furnish an A.C. component with the frequency ω_3, then instead of $-A_1 \cos \omega_h t$ we must use the term $A_2 \cos 2\omega_h t$ or $-A_3 \cos 3\omega_h t$, etc. Instead of A_1 we then have everywhere A_2 or A_3, etc., so that the conversion amplification becomes

$$a_c = \frac{{}^1/_2 A_2}{\dfrac{1}{R_0} + A_0} \quad \text{or} \quad a_c = \frac{{}^1/_2 A_3}{\dfrac{1}{R_0} + A_0} \text{ etc.} \quad . \; . \; . \; (12)$$

Now for very small values of b the coefficients A_1, A_2, A_3, etc. of the series (1) all approach the same value, namely $b.s_{\max}/\pi$, which shows that with sufficiently small values of b the application of this manner of frequency transformation does not mean any depreciation in conversion amplification in the case of a diode. For use at very high frequencies this property is especially important. Since the coefficients A_n approach the limiting value mentioned with increasing value of n more slowly upon decreasing b, in practice in frequency transformation by means of a higher harmonic than the fourth or fifth an appreciable decrease in conversion amplification will, however, occur.

Practical realization of the connections

As described in the above the D.C. voltage V applied to the diode must be about equal to the amplitude E_h of the oscillator voltage. In practice the voltage V is formed by allowing the diode current to flow through a leakage resistance. Connections for this are indicated in *fig. 5*. The high-

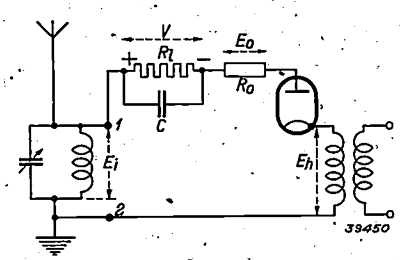


Fig. 5. Connections of a diode as frequency-changing valve.

frequency voltage is taken from an oscillator circuit coupled with an aerial, while the oscillator voltage is induced in a coil in the cathode connections. The condenser C has such a large capacity that it represents practically no impedance for the alternating currents occurring. The D.C. component of the diode current, however, flows through the leakage resistance R_l and causes the occurrence of the D.C. voltage V therein. With a sufficiently large value of R_l, $V-V_0$ will be almost equal to the amplitude of the A.C. voltage E_h.

The part of a period of the oscillator voltage during which current flows in the diode, thus the angle b of fig. 4, can easily be determined approximately from the size of the leakage resistance and the properties of the diode. For this purpose the diode characteristic is approximated by a straight line. The reciprocal of the slope of this line is called the statically measured internal resistance R_d of the diode (*fig. 6*). If we call the negative voltage at which current just begins to flow in

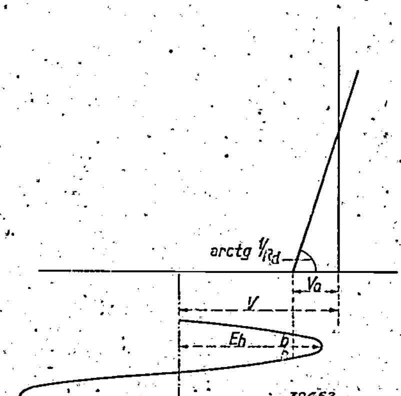


Fig. 6. Approximation of the diode characteristic by a straight line.

the diode, V_0, then during the short time that current flows through the diode that current is given by

$$\frac{E_h \sin \omega_h t - (V - V_0)}{R_d}.$$

The charge flowing through the diode during one period is

$$\int_{(^1/_2\pi-b)/\omega}^{(^1/_2\pi+b)/\omega} \frac{E_h \sin\omega_h t - (V-V_0)}{R_d}\,dt = \frac{2}{\omega_h}\left(\frac{E_h}{R_d}\sin b + \frac{V-V_0}{R_d}\,b\right).$$

This charge is equal to that which flows through the leakage resistance R_l in a period, namely $\frac{V}{R_l}\frac{2\pi}{\omega_h}$; therefore

$$\frac{E_h}{R_d}\sin b - \frac{V-V_0}{R_d}\,b = \frac{V}{R_l}\pi.$$

If one considers that $(V-V_0)/E_h = \cos b$, after some work one finds

$$\operatorname{tg} b - b = \frac{V}{V-V_0}\cdot\pi\,\frac{R_d}{R_l} \quad \ldots \quad (13)$$

If, as is usually the case, one sets $V_0 \ll V$, this equation may be simplified to

$$\operatorname{tg} b - b = \pi\,\frac{R_d}{R_l} \quad \ldots \quad (14)$$

If b is known, the value of $V-V_0$ is found from the condition $V-V_0 = E_h \cos b$, and with the help of (14) it may be seen that upon increasing the value of R_l the D.C. voltage $V-V_0$ actually does lie closer and closer to E_h.

In the practical application of the connections the input and output damping of the frequency-changing stage are important as well as the conversion amplification. The input damping is given by the resistance of the part of the connection to the right of points *1* and *2* in fig. 5 for the frequency ω_i. The latter is again found from equation (4) by taking this time from the series for $s(t)$ the terms which account for the occurrence of the frequency ω_i. These are again the term A_0, which, multiplied in $E_i \sin \omega_i t$, gives $A_0 E_i \sin \omega_i t$, and the term $-A_1 \cos \omega_h t$ which multiplied by $E_0 \sin \omega_0 t$ gives

$$-A_1 E_0 \sin \omega_0 t \cdot \cos \omega_h t = \\ = {}^1/_2 A_1 E_0 \left\} \sin (\omega_h - \omega_0)t - \sin (\omega_h + \omega_0)t \right\{.$$

Since $\omega_i = \omega_h - \omega_0$, the current component with the frequency ω_i is

$$i_i \sin \omega_i t = A_0 E_i \sin \omega_i t + {}^1/_2 A_1 E_0 \sin \omega_i t.$$

According to equation (6)

$$E_0 = -E_i\,\frac{{}^1/_2 A_1}{\dfrac{1}{R_0} + A_0}.$$

For the resistance of the part of the connections to the right of points *1* and *2* for the frequency ω_i we now find:

$$r_i = \frac{E_i}{i_i} = \frac{1/R_0 + A_0}{A_0^2 - {}^1/_4 A_1^2 + \dfrac{A_0}{R_0}}.$$

Now according to (8), $A_0 = 1/r_d$, while according to (9) with the small values of b occurring in practice A_1 approaches $2/r_d$. If we substitute this in the last equation we find that:

$$r_i = r_d + R_0 \quad \ldots \quad (15)$$

or, in words: the resistance with which the frequency-changing circuit damps the input circuit is equal to the sum of the internal resistance of the diode and the resistance of the intermediate-frequency circuit (or of the primary side of the intermediate-frequency transformer) for the output frequency. The degree to which the input circuit is damped in these connections thus depends upon the construction of the intermediate-frequency circuits.

Since the input voltage E_i and the output voltage E_0 are in series in the connections, the same reasoning which was followed above for the determination of the damping of the input circuit can also be applied for the determination of the damping which the connections exert on the output circuit, thus on the intermediate-frequency circuit or transformer. This resistance is therefore again equal to the sum of the internal resistance of the diode and the resistance which the input circuit represents for the frequency ω_i. If we call this last resistance R_i, then according to the preceding the resistance with which the frequency-changing connections damp the output circuit is given by

$$r_0 = r_d + R_i \quad \ldots \quad (16)$$

A change in the input circuit, for instance a shift in its tuning over a given wave-length region, will in general cause a change in the damping of the first intermediate-frequency circuit of the set in question, and thus also a change in the selectivity.

The value of r_0 given by equation (16) is identical with the internal resistance of the frequency-changing connections for intermediate frequency, which has previously been mentioned in connection with the possibility of obtaining a conversion amplification greater than unity by connecting R_0 *via* a transformer.

The internal resistance of the diode according to (8) and (2) is given by

$$r_d = \frac{1}{A_0} = \frac{2\pi}{b \cdot s_{\max}}.$$

A large value of r_d, and thus also of r_i and r_0, can apparently be attained by making b small. According to equation (14) a large value of the leakage resistance R_l is necessary for this purpose. Thus while the use of a large leakage resistance is favourable in connection with the damping of the circuit connected with it, this cannot be carried too far. From equation (11) it may be concluded that a very large value of r_d gives a decrease of the conversion amplification reached. In practice, therefore, in choosing the leakage resistance a compromise should be made between the conversion amplification obtained and the damping of the circuits in connection.

The theory of frequency transformation given above is actually only more or less exact at frequencies which are not so high that the transit time of the electrons in a diode is important. Wave lengths shorter than 1 m, however, no longer satisfy this condition. Due to the inertia of the electrons a phase shift occurs between an A.C. voltage applied to the diode and the alternating current caused hereby. Thus for example the periodic variation in slope caused by the oscillator voltage will no longer be in phase with the oscillator voltage, and instead of fig. 3, a variation of slope will occur which is unsymmetrical with respect to π, 3π, etc. Fig. 3 can then no longer be approximated by a simple set of triangles as sketched in fig. 4. For all these reasons it is evident that an exact theory of frequency transformation by means of a diode at very high frequencies leads to very complicated calculations. We shall therefore confine ourselves in this article to the reproduction of several results which were obtained with this type of frequency-changing valve on decimetre waves (see the following).

Push-pull frequency-changing stages [4])

In our considerations of diode frequency changing we dealt always with connection with a single diode. Now it has already been shown previously in this periodical that in the solution of the problem of voltage-amplification at very high frequencies, advantages can be obtained when use is made of A.C. voltages which are balanced with respect to earth. In this connection for example the high-frequency amplifier valve EFF 50 was designed as a push-pull amplifier valve [5]).

Continuing on this line of thought the frequency-changing valve for decimetre waves was now also provided with two diodes, whereby frequency transformation of A.C. voltages which are balanced with respect to earth is possible.

Two A.C. voltages in opposite phase are thus applied to these diodes. If we call the high-frequency voltage on one diode $^1/_2\,E_i \sin \omega_i t$, then the high-frequency voltage on the other diode is $-^1/_2 E_i \sin\omega_i t$. As concerns the oscillator voltage, there are now two possibilities. The oscillator voltage can be fed to the two diodes either in phase or in opposite phases. Connections are given in *fig. 7a* and *b* for

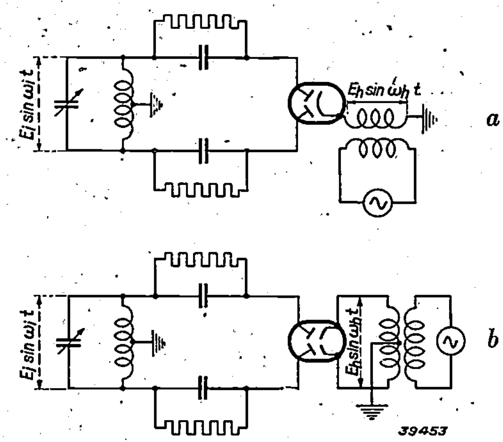


Fig. 7. Diagrams showing the principle of a push-pull frequency-changing stage.
a) Upon feeding the oscillator voltage to the two diodes in the same phase.
b) Upon feeding the oscillator voltage to the two diodes in opposite phase.

these two possibilities. In the first case a single common cathode for the two diodes is found to be sufficient, while in the second case each diode is provided with a separate cathode. In the connections according to fig. 7a intermediate-frequency voltages are obtained on both diodes which are in opposite phase with respect to earth, while in the connections according to fig. 7b, where in addition to the input voltage the oscillator voltage is also fed to the two valves in opposite phase, two intermediate-frequency voltages are obtained which have the same phase with respect to earth. This difference between the connections of fig. 7a and b should be kept in mind in adding the intermediate-frequency circuit to the connections.

Frequency changer for decimetre waves

We shall now give a description of a frequency-changing valve with indirectly heated cathode, suitable for the reception of decimetre waves. In this valve two diodes in push-pull connection are used, while at the same time a triode for exciting the oscillator voltage is also built in. (For the excitation of the oscillator voltage two triodes in push-pull connection could also be used. In this

[4]) See M. J. O. Strutt, Moderne Kurzwellenempfangstechnik, Springer, Berlin 1939, p. 182.

[5]) Philips techn. Rev. 5, 172, 1940.

description, however, we shall confine ourselves to a type with one triode).

For the sake of the simpler construction and connections the principle of fig. 7*a* was chosen. *Fig. 8* shows a laboratory model of this valve as well as a photograph of the interior of the valve. The complete connections of the valve as a frequency changer are shown in *fig. 9*. The oscillation circuit *I* is the high-frequency input circuit on which the high-frequency A.C. voltage acts which is fed to the two diodes *via* the condensers C and C_1. The oscillation circuit is the circuit *II*, while *III* is the intermediate-frequency circuit. The triode part of the valve, together with the oscillator circuit *II* forms an oscillator in the so-called three-point connection in which the division of voltage takes place by means of the internal capacities of the valve. A choking coil S_2 is included in the cathode connection, on which an A.C. voltage with the oscillator frequency occurs. Since the cathodes of the triode part and of the diode part are connected to each other, the desired oscillation voltage occurs in this way on the cathode of the two diodes. Choking coils (S_2' and S_2'') are also included in the two heating current lines. The intention is hereby to prevent the capacity between cathode and heating filament becoming connected in parallel with S_2, which might cause a decrease in the impedance between cathode and earth, or lead to disturbing resonance phenomena. In order to eliminate the influence of any fluctuations in the capacity between cathode and filament a condenser is introduced between the two.

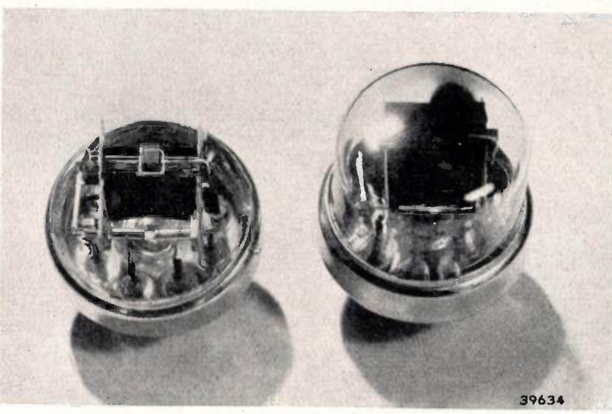


Fig. 8. Laboratory model of the diode frequency changer. Left interior view; right external appearance. The shielding which surrounds the triode is partly removed to show the interior.

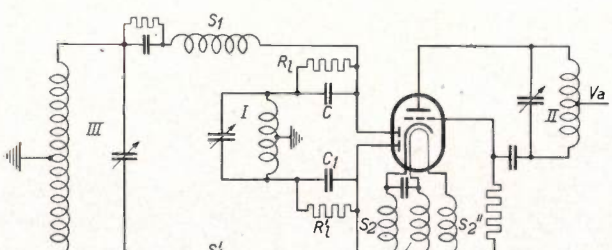


Fig. 9. Connections of the diode frequency changer, in which the oscillator voltage is fed to the two diodes in the same phase.

The two intermediate-frequency voltages in opposite phase which are delivered by the diodes are fed to the intermediate-frequency circuit *III via* the choking coils S_1 and S_1'. In series with these coils two large condensers are connected which prevent the leakage resistances Rl and Rl' for direct current from being short circuited *via* these choking coils and the coils of the circuits *I* and *III*. The self-induction of S_1 and S_1' is chosen so great that they possess a high impedance for the frequency of the high-frequency voltage to be received, while for the much lower intermediate frequency the impedance of S_1 and S_1' is slight. The capacity of the condensers C and C_1 is kept so small that the condensers have a high impedance for intermediate-frequency A.C. voltages, while for the much higher frequency of the high-frequency A.C. voltage the impedance of C and C_1 is slight. In this way the oscillation circuits *I* and *III* are prevented from affecting each other. The circuit *III* can naturally also be formed by the primary circuit of an intermediate-frequency transformer.

At wave lengths shorter than 1 metre the oscillation circuits *I* and *II* cannot as a rule be composed of a coil and a condenser. A system of parallel conductors (a Lecher system) is usually used for this purpose, which is provided with a sliding short-circuit bridge [6]). The length of the conductors calculated up to the short circuit must then be adjusted approximately equal to an odd number of quarter wave lengths for the desired tuning frequency.

By means of different devices the diode part as well as the triode part of this valve have been given the best possible properties for use on decimetre waves. In the first place the distances between the electrodes in both parts have been made as small as possible so that the transit time of the electrons is limited to a minimum. In the case of the two diodes the distance between anode and cathode amounts only to 0.15 mm. Furthermore the connections of the electrodes to the pins are made as

[6]) Philips techn. Rev. **6**, 241, 1941.

short as possible. The triode part may be used as oscillator down to a wave length of 37 cm. Upon the use of for instance the fourth harmonic of the oscillator voltage, therefore, frequency transformation can still be obtained at a wave length of about 9 cm.

Special attention is paid to the damping which the valve exerts on the oscillation circuits connected to it. In addition to the damping which occurs when a given current flows in the diode, which we may call the active damping, and which in the first part of this article we have already calculated, there is also another source of damping present. Between anode and cathode of each diode there is a certain capacity; these capacities are connected in series with the input oscillation circuit. In series with these capacites is the resistance of the connection pins which may cause a considerable damping of the circuit in question.

The ordinary material for the construction of connection pins is chrome-iron. Due to the skin-effect pins made of this material take on a rather high resistance on decimetre waves. This may for instance, for the part of a pin which runs through the glass wall, amount to as much as several ohms. In order to diminish these resistances the whole of the chrome-iron pin is copper-plated. The part which projects outside the glass is covered in addition with a layer of silver. It has been found that the damping of the circuits connected to them is hereby appreciably decreased. It must still be mentioned that the short wave properties of the triode part can be considerably improved by connecting the anode and (or) grid to two instead of one connection pin.

As has already been calculated, the conversion amplification which can be attained when the intermediate-frequency circuit is connected directly and not *via* a transformer is always less than unity. With the valve described measurements of the conversion amplification and of the input damping in connections like those of fig. 9 were carried out at a wave length of the input signal of 1.2 metres. The oscillator circuit was here tuned to wave lengths of 2.4, 3.6 and 4.8 metres, so that the frequency transformation took place by means of the 2nd, 3rd and 4th harmonic, respectively, of the oscillator voltage. If a leakage resistance R_l of 50 kilo-ohms was used and the intermediate-frequency circuit was tuned to a frequency of 30 megacycles/sec (wave length 10 m), conversion amplifications were measured of 0.63, 0.55 and 0.40 [7]), respectively. When the intermediate-frequency circuit was tuned to a lower frequency, namely 10 megacycles/sec, it amounted to about 7 kilo-ohms.

Fig. 10 shows an arrangement for measuring the

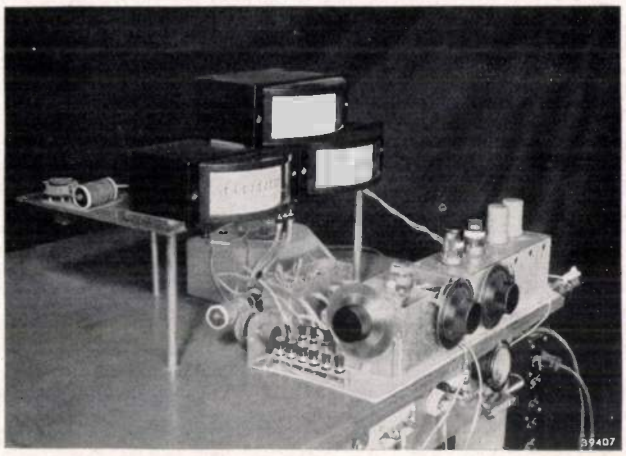


Fig. 10. Arrangement for the measurement of the conversion amplification.

conversion amplification. The scheme of this arrangement is represented in *fig. 11*.

In the constructions of connection according to *fig. 9* or *fig. 11* special attention must be paid to the symmetry of input and output circuits and the lines connected to them with respect to earth. Only in the case of such symmetry is an absolute independence of the oscillator circuit *II* and the other two circuits assured. In order to make this clear several capacities present in the connections are indicated in *fig. 12*. C_{d_1} and C_{d_2} are the capacities of the two diodes, C_1, C_1', C_2 and C_2' are the capacities of the wiring toward earth, while C_{ak} and C_{gk} are the capacities of anode and grid, respectively, of the triode part toward earth. If C_1 and C_2 are equal to C_1' and C_2' then, since $C_{d_1} = C_{d_2}$, in the presence of an A.C. voltage between the extremities of oscillation circuit *I*, no A.C. voltage with the frequency ω_i will occur between the cathode of the frequency changer and earth. The high-frequency voltage will therefore have no influence on the oscillator circuit *II*. Conversely, in the presence of an oscillator voltage on the cathode, no A.C. voltage with the oscillator frequency will be able to occur across the oscillation circuit *I*. The circuit *I* and *II* are thus entirely independent of each other. If on the contrary C_1 and C_2 are not equal to C_1' and C_2', the foregoing is not the case. In the presence of an A.C. voltage between the extremities of circuit *I* there will also be an A.C. voltage on the cathode which also causes the occurrence of an A.C. voltage on the oscillator circuit *II* *via* the capacities C_{ak} and C_{gk}.

[7]) Philips techn. Rev. 5, 172 and 257, 1940.

Conversely, an oscillator voltage present on the cathode can then also cause an A.C. voltage over the oscillation circuit *I*. The result is that a change in tuning of one of the oscillation circuits *I* or *II* also affects the tuning frequency of the other circuit, which makes the adjustment very much more difficult. A second result is that the oscillator circuit *II* is damped due to the presence of circuit *I*, whereby the oscillation of the triode part sometimes becomes more difficult.

voltage E_0 with the frequency ω_0 acts between these two anodes, E_i must be measured with the help of a selective amplifier (*B* in fig. 11) which only amplifies an A.C. voltage with the frequency ω_i. The measurement is made still more difficult by the fact that the occurrence of an A.C. voltage with the frequency ω_h on the oscillation circuit *I* can practically never be entirely avoided. On very short waves and in the case of frequency transformation by means of the first harmonic of the oscillator voltage, the difference in tuning frequency between the circuits *I* and *II* is only slight. This fact sets high requirements on the selectivity of the measuring amplifier in question.

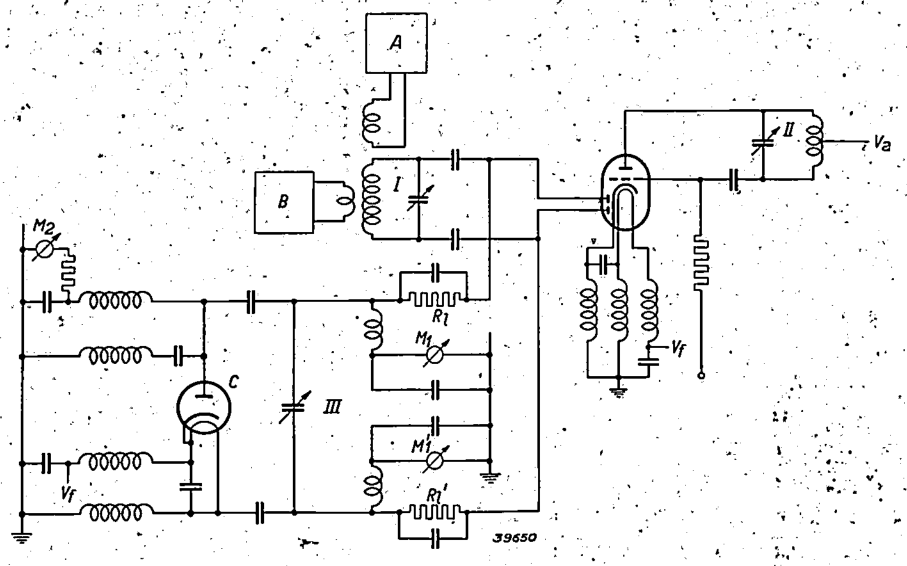


Fig. 11. Diagram of the arrangement for measuring conversion amplifications. *I* = high-frequency circuit, *II* = oscillator circuit, *III* = intermediate-frequency circuit, *A* = transmitter, *B* = selective amplifier for measuring the high-frequency voltage between the diode anodes, *C* = diode for measuring the intermediate-frequency voltage, M_1 and M'_1 = instruments for measuring the direct current in the grid resistances R_l and R_l'. M_2 instrument for measuring the direct current in the diode *C*.

The necessity of the symmetrical construction of the high-frequency circuit *I*, which was demonstrated above, is also valid of course for the intermediate frequency circuit *III* (fig. 9). Thus for example the choking coils S_1 and S_1' should be as similar as possible, while they should be mounted in the same way in relation to metal screens connected to earth.

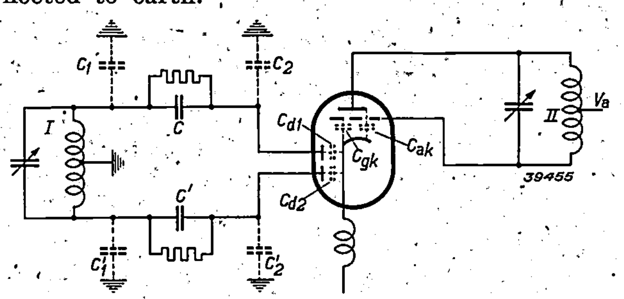


Fig. 12. Diagram of the diode frequency changer indicating the different capacities which play a part in the connections.

The measurement of the high-frequency voltage E_i which acts between the two diode anodes should be carried out with great care on the measurements here in question. Since in addition to the A.C. voltage E_i with the frequency ω_i the A.C.

This amplifier is calibrated by determining the proportionality factor, with the oscillator switched off, between the high-frequency A.C. voltage on the diodes and the deviation of a measuring instrument connected to the amplifier. The voltage on the diodes is hereby determined from the direct current flowing in the leakage resistances.

A complete series of valves for reception of decimetre waves

With the development of the frequency-changing valve described, an entire series of electronic valves for the reception of decimetre waves has been completed. The series consists of the following valves: the valve EFF 50 as high-frequency amplifier valve, the valve EF 51 as intermediate-frequency valve and the above described valve as frequency changer. A detailed description of the first two valves has already been given in this periodical [8]).

The push-pull amplifier valve EFF 50 can be used as high-frequency amplifier approximately down to a wave length of 50 cm. To this wave length

[8]) Philips techn. Rev. 5, 172 resp. 357, 1940.

365

therefore the complete series mentioned may be used. The wave-length of the intermediate-frequency oscillations may amount to from several metres to approximately one hundred metres. For wave lengths shorter than 50 cm the high-frequency amplification with the valve EFF 50 is less effective. In such a case the frequency-changing valve described can be used as first valve in the receiving set. As was mentioned above, this valve can be used down to a wave length of about 9 cm.

A diode frequency changer for battery supply

In addition to the valve shown in fig. 8, a short-wave frequency changer for battery supply has also been developed. Due to the large heating current consumption which is always connected with the use of an indirectly heated cathode, the valve described is less suitable for battery supply. The battery frequency changer is therefore of the directly heated type. The heating current for this valve amounts to only 200 mA with a voltage of 1.2 volts, compared with 300 mA with a voltage of 6.3 volts in the case of the indirectly heated type. Due to the fact that the filaments are strung in the shape of a W, the distance between the cathode and the other electrodes cannot be made as small as in the case of the indirectly heated cathode. As a result the triode part can only be used as oscillator down to a wave length of 80 cm. The two diodes, however, still work at considerably shorter wave lengths.

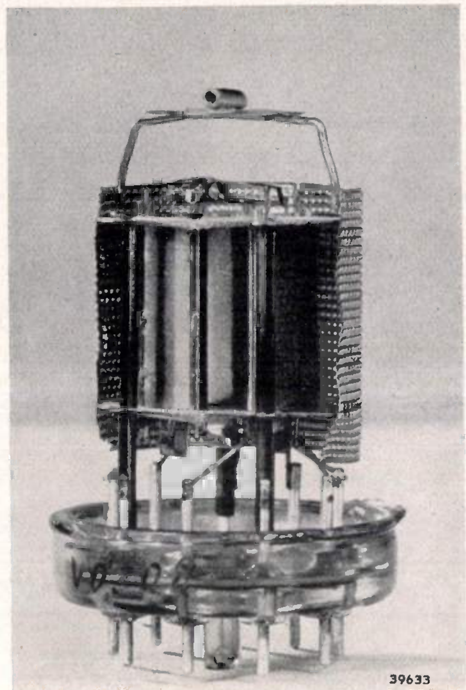


Fig. 13. Interior of the diode frequency changer for battery supply. Part of the shielding which surrounds the whole has been removed.

Fig. 13 is a photograph of the internal construction of a laboratory model of the diode frequency changer for battery supply. In the photograph part of the metal shield which surrounds the two diodes and the triode is cut away.

THE FLICKERING OF SOURCES OF ELECTRIC LIGHT

by P. J. BOUMA.

621.32

The variation of the light intensity during half a period of the alternating current (1/100 sec) is given for different technical light sources. The degree is studied to which this change in light intensity is observed as a flicker, a distinction being made between two cases, namely the observation of stationary and that of moving objects. In the first case it is found from measurements of the critical flicker frequency carried out under different circumstances that in the case of light sources which burn on an A.C. main of 50 c/s the flickering of stationary objects can only be observed at extremely high brightnesses. In the second case the varying light intensity may much more easily cause trouble due to the occurrence of stroboscopic effects. The contrasts which are observed in the field of vision under such circumstances are calculated for different cases.

Introduction

When a source of electric light (incandescent filament lamp, gas-discharge lamp) is connected to an A.C. voltage, voltage and current and consequently the amount of light radiated, will exhibit periodic variations. Under certain conditions the eye is capable of observing the results of these periodic changes. The lamp in that case is said to flicker[1]).

At what frequency does this phenomenon of flicker appear? In our discussion we assume that the lamp is always connected to an A.C. voltage of 50 c/s. If the oscillation time of 1/50 sec is divided into two halves of 1/100 sec, the same voltage variation takes place during each half, except for the sign. Most sources of light have the property that the current also shows the same variation in the two halves of the period except for the sign, and that the light intensity also varies in the same way in the two halves of the period. From this it follows that the number of periods of the flicker phenomenon will in general amount, not to 50, but to 100 c/s.

Exceptions to this rule occur when the light source does not behave in the same way in the two phases of the A.C. voltage. This is the case when the lamp has a rectifying action (the light of mercury rectifiers), or when for instance in a gas-discharge lamp the two cathodes are not alike. In these cases flicker phenomena occur at 50 c/s. Since the latter frequency of flicker is much more disturbing than 100 c/s, an attempt will always be made to avoid such lack of symmetry.

The variation of the light of several sources of light during the period of 1/100 sec.

In *figs. 1-5* may be seen the variation of the light during half a period of the alternating current (1/100 sec) for several types of light sources which were measured in this laboratory. The phase angle φ (expressed in degrees) is plotted along the abscissa. As zero point of the measurement in each case the moment was chosen at which the light radiation is a maximum. As ordinate the ratio was chosen of the light intensity I at a given moment to the maximum light intensity I_0.

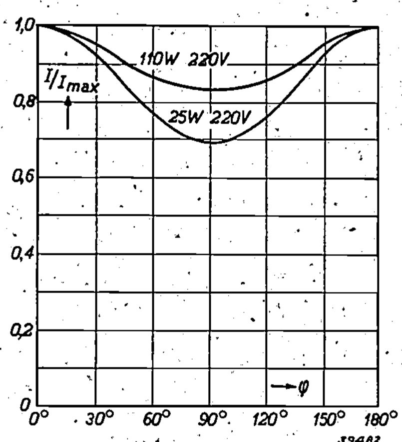


Fig. 1. Light variation of electric lamps during one half of a period (1/100 sec), in which the phase angle φ of the voltage increases from 0 to 180°.

Fig. 1 refers to gas-filled electric lamps. The variation here is practically sinusoidal. The magnitude of the ripple of 100 c/s depends very closely upon the heat capacity of the filament; it may be seen that low-power lamps give greater variations than larger types.

In *fig. 2* are shown the curves of the light variation for different mercury lamps [2, 3]). It may be seen that at very high mercury pressures (SP 500 W) the part of the period in which very little light is radiated increases rapidly.

Fig. 3 refers to the blended light lamp ML 500 [4]). If the light intensity is measured in the direction

[1]) Other flicker phenomena than the one here mentioned, such as the flicker of luminous flames, fall outside the scope of this article.

[2]) Philips techn. Rev. 1, 129, 1936.

[3]) Philips techn. Rev. 2, 165, 1937.

[4]) Philips techn. Rev. 5, 341, 1940.

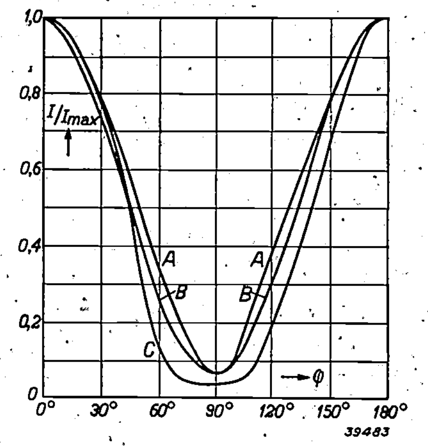


Fig. 2. Light variation of mercury lamps during half a period (1/100 sec) of the voltage. *A*) Low-pressure mercury lamp. *B*) Type HP 300. *C*) Type SP 500 W.

of the axis of the lamp, a much smaller variation is found than in a direction perpendicular to it (curves *A* and *B*). This is due to the fact that in the former direction the contribution of the ordinary electric lamps is relatively much larger than in the latter direction. Curve *C* gives the variation of the total light flux.

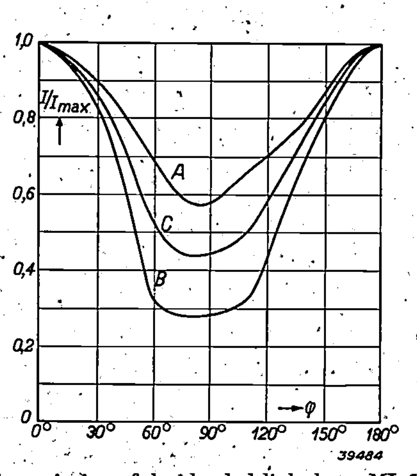


Fig. 3. Light variation of the blended-light lamp ML 500 during half a period (1/100 sec.).
A) Measured in the direction of the axis of the lamp.
B) Measured perpendicular to this direction.
C) Average light variation.

In *fig. 4* may be seen the variation for low-pressure mercury lamps with luminescence (TL 100 [5])). If we compare these curves with fig. 2 (curve *A*), we see that the light intensity of the lamp TL 100 varies much less: during the part of the period in which the mercury discharge itself emits no radiation at all, the luminescent substance emits, and thus partially fills up the sharp minimum of fig. 2. The extent of this filling up depends very much upon the choice of luminescent substances.

Fig. 5 gives the light variation for the sodium lamp SO 650 [6]). The striking unsymmetrical form of this curve is a result of the very low sodium pressure at which the lamp works in order to obtain high efficiency. When the current is large, a large part of the sodium vapour is ionized and therefore can no longer emit the yellow lines, so that the latter are then only produced by a thin layer at the wall of the discharge tube, while inside it neon light is emitted. The formation and disappearance of this film on the wall does not take place symmetrically within a period.

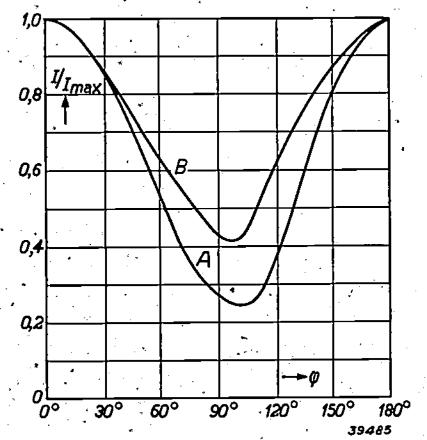


Fig. 4. Light variation of luminescence lamps TL 100 during half a period (1/100 sec) of the voltage. *A*) Daylight lamp. *B*) Lamp with colour temperature of about 3 500 ° (warm white).

The light variation of a light source can also be determined by developing the curves of figs. 1-5 into Fourier series:

$$I = \overline{I}\,\{1 + a_1 \sin(\omega t + \varphi_1) + a_2 \sin(2\omega t + \varphi_2) + \ldots\} \; . \quad (1)$$

In this series I is the light intensity at a given moment t; $\overline{I}$ the average light intensity; $a_1, a_2, \ldots$ the coefficients of the different Fourier components; $\omega = 2\pi$ times the fundamental frequency

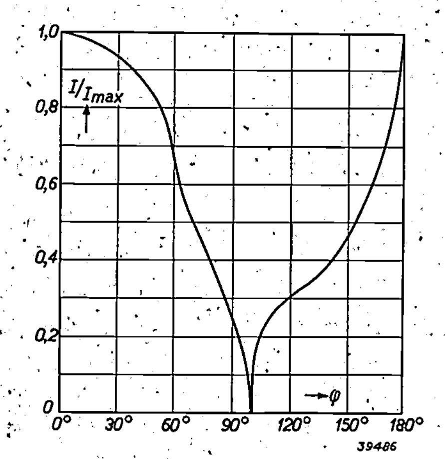


Fig. 5. Light variation of the sodium lamp SO 650 during half a period (1/100 sec) of the voltage.

[5]) Philips techn. Rev. 4, 337, 1939; 6, 65, 1941.

[6]) Philips techn. Rev. 2, 353, 1937.

(here 1/100 sec) and φ_1, φ_2, . . . certain phase shifts. For the curves of figs. 1-5 we then find the values for a_1, a_2, a_3 given in *table I*.

Table I

Kind of lamp	a_1	a_2	a_3
Electric lamp 110 W 220 V	0.0923	0.007	0.002
Electric lamp 75 W 220 V*)	0.123	0.003	—
Electric lamp 25 W 220 V*)	0.179	0.016	—
Low-pressure mercury lamp*)	0.82	0.078	—
Mercury lamp HO 1600	0.750	0.021	0.017
Mercury lamp HP 300	0.890	0.021	0.006
Mercury lamp SP 500 W	1.126	0.132	0.047
Blended light lamp ML 500			
A) in the direction of the axis	0.244	0.016	0.024
B) perpendicular to *A* direction	0.622	0.045	0.055
C) average	0.433	0.030	0.040
Mercury lamp TL 100			
Daylight	0.569	0.048	0.020
Warm white	0.391	0.034	0.034
Sodium lamp SO 650	0.710	0.068	0.098
Sodium lamp SO 400	0.599	0.072	0.098

*) E. G. Andresen, Das Licht 7, 235, 1937.

The flickering of stationary objects

Until now we haven spoken only of the purely physical phenomenon of the light variation within the period. As was mentioned in the introduction, one may speak of flickering only when the eye can observe the results of this variation.

These results may be divided into two large groups:

a) When a non-moving object is observed, under certain conditions the variations of the brightness of the object can be observed directly.
b) When moving objects fall within the field of vision, stroboscopic effects can be observed: one sees for instance instead of a continuously moving rod, a whole series of separate stationary images of the rod.

We shall first deal with the first group of phenomena.

When the conditions (brightness, frequency of the flicker, magnitude of a_1, a_2, a_3 . . . ,etc.) are so chosen that a stationary object is seen to flicker, and the frequency is then steadily increased, the flickering will finally be seen to stop. The frequency at which the flicker phenomenon stops is called the critical flicker frequency.

This critical frequency depends upon various circumstances.

In the first place upon the form of the light variation during the period. From numerous experiments with different vibrational forms it has been determined that only the quantity a_1 of formula (1) is important here: the higher harmonics (overtones) have no influence on the values of the critical frequency, unless one of the overtones is several times as intense as the fundamental. According to table I the latter is not the case for any of the commonests light sources: the strongest overtones occur with the sodium lamp, but even here they do not amount to more than about 15 per cent of the fundamental.

In connection with this property we may in the further investigation confine ourselves to a single definite vibration form. We have chosen the following light variation for this purpose.

The brightness of the object amounts to B_1 during one half of the period, during the other half it is B_2; the average brightness is therefore $\bar{B} = \frac{1}{2}(B_1 + B_2)$.

In a preliminary series of observations B_2 was equal to zero. A circular white disc was chosen as flickering object. It could be observed within different angles of vision and its centre was always focussed on. The surroundings of the disc were dark. *Fig. 6* shows how under these conditions the critical frequency n depends upon the average brightness B and the magnitude of the angle of vision.

The following was noted:

1) n increases almost linearly with log B in the brightness region from 0.5 to 200 cp/m², which is of greatest practical importance;

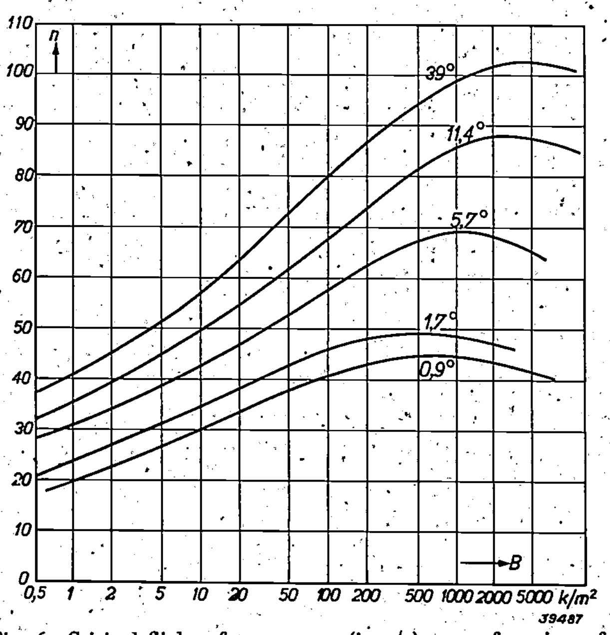


Fig. 6. Critical flicker frequency n (in c/s) as a function of the average brightness B, with circular fields of vision observed within angles from 0.9 to 39°. The brightness was equal to $2B$ in the first half of a period and zero in the second half. The surroundings of the field of vision were dark.

2) as the field of vision is made greater a larger value of n is found.

The first property is known under the name of the Ferry-Porter law. Only at very high and very low brightnesses do deviations from the law clearly occur. As to the slope of the straight sections of the curves drawn in fig. 6, there is surprisingly good agreement among the different authors. Thus for example in the case of very careful and detailed observations in Amerca [7]), for an angle of vision of 2°, slopes were found which varied between 10 and 11, while in fig. 6 we find for 1.7° and between 0.5 and 200 cp/m² an average slope of 10.8.

The increase of n with the angle of vision is connected with the fact that the peripheral parts of the eye are more sensitive to rapid light changes than the centre of the eye. Furthermore we see from fig. 6 that we will practically never have trouble with the flickering of stationary objects when the light varies at 100 c/s. Even with the largest field of vision brightnesses of more than 1 000 cp/m² are necessary to be able to observe the flicker. The illuminated objects will never reach such brightnesses (for a pure white object an illumination intensity of about 4 000 lux would be required). The light sources themselves will, however, be able to exhibit such high brightnesses, but they are always seen within much smaller angles of vision, so that here also the critical frequency of $n = 100$ is never reached at the highest brightnesses.

In a second series of observations B_1 as well as B_2 was varied. For the rest the observations were carried out in the same way with an angle of vision of 39°. The results were worked out in the following way. For a light variation as described above we find upon development into Fourier components that

$$a_1 = \frac{4}{\pi}\frac{B_1 - B_2}{B_1 + B_2}.$$

Each observation thus provided us with a corresponding set of values for B, a_1 and n. The relation thus found between B, a_1 and n has the advantage that the results may be applied directly to a light variation of a different vibrational form. For $B_2 = 0$, the first case investigated (fig. 6), we obtain the maximum value $4/\pi$ which a_1 can take on.

We here find the same result as from fig. 6, namely that a flicker with 100 c/s can only be observed at extremely high brightnesses and only for values of a_1 in the neighbourhood of 1. The situation would be quite different if the light varied at 50 c/s. If we then used for example a gas-discharge lamp with $a_1 = 0.50$, the flicker would already be visible at a brightness of abour 10 cp/m², *i.e.* when a piece of white paper was illuminated with about 40 lux, while if an electric lamp with $a_1 = 0.20$ were used, illumination intensities higher than about 150 lux would not be permissible.

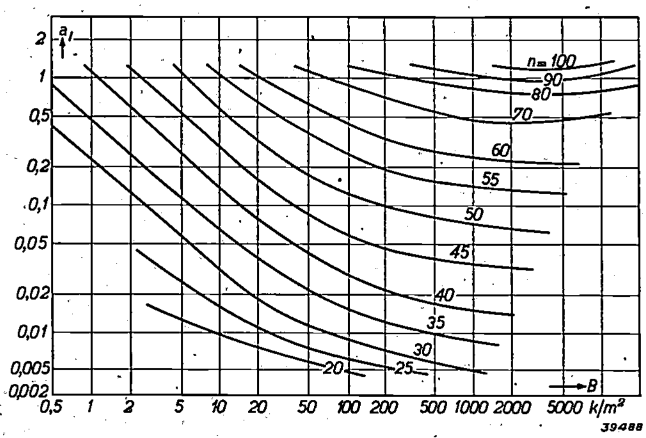


Fig. 7. The brightness of a circular disc in dark surroundings, observed within an angle of vision of 39°, varies with the time according to the formula: $B(t) = B(1 + a_1 \sin 2\pi tn)$. The figure gives the values of a_1 (as a function of B and n) at which the flickering is just not visible.

The lowest value of a_1, which we can observe is found to be about 0.005 at frequencies of 20 to 30 c/s and brightnesses of at least 100 to 1 000 cp/m².

In the investigations in the regions of very low values of a_1 the surprising fact is encountered that there are two critical frequencies; only in the region between these two frequencies can the flicker be observed. Upon further consideration this property is obvious: if the light intensity is made to vary extremely slowly, we will no longer be able to observe the variation. With high values of a_1 and of the "upper critical frequency" n the "lower critical frequency" lies much lower than n; in the region of low values of a_1, however, the two critical frequencies lie closer and closer together and finally coincide at $a_1 \approx 0.005$, so that the flicker region then disappears entirely.

The flickering of moving objects

In the foregoing it was found that with a variation of the light intensity at 100 c/s there is no danger of the occurrence of flicker with stationary objects. On the other hand it is well known that at the same frequency moving objects can be seen to flicker, in other words stroboscopic effects can be observed.

This phenomenon is considerably more compli-

[7]) S. Hecht and C. D. Verrijp, Proc. Nat. Acad. Sci. 19, 522, 1933.

cated than that of the flicker of stationary objects; this complexity is due especially to the large number of variables which must in these cases be taken into account. In the theory of the phenomenon which will be developed below we shall in many respects be compelled to submit to restrictions. We begin with the common method of determining "whether or not a light source flickers", *i.e.* we take a light coloured, preferably glossy rod (pencil, scissors) and move it rapidly back and forth to see if we can observe separate images. If this process is put into a somewhat more mathematical form, it runs as follows. Move a rod of thickness d over a uniformly illuminated background with the velocity v. Let the rod take on a brightness B and the background the brightness B_0 ($B > B_0$) with an illumination of 1 lux. Rod and background are both illuminated by a light-source of which the variation of the light intensity during the period T is known. The quantities required are the contrasts which are seen to appear in the field of vision.

As to the variation of the light intensity, we shall confine ourselves to the types sketched in *fig. 8*. The following times are of importance for the further discussion of the problem.

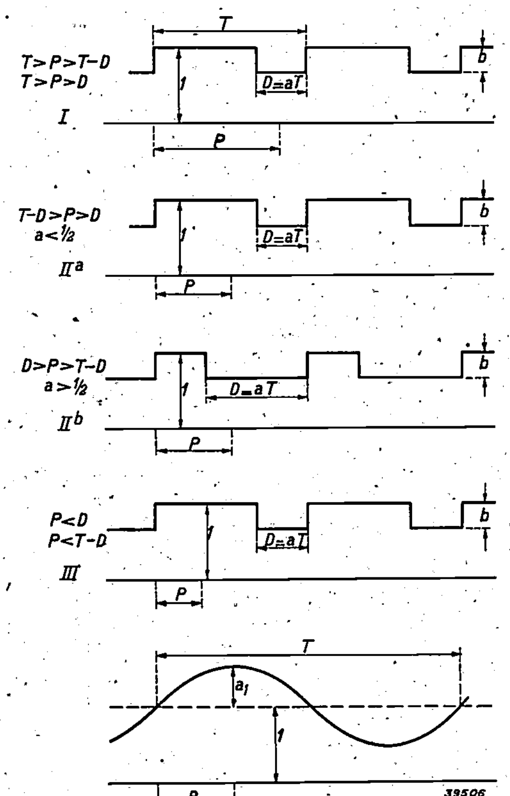


Fig. 8. Different forms of the light variation for which the contrasts occurring in the field of vision upon observation of moving objects are calculated.
T, D and P are the time of vibration, the dark period and the time of passage (time in which the moving object entirely passes a given point), respectively.

The oscillation time T, the "dark period" $D = aT$ and the "time of passage" $P = d/v$, *i.e.* the time needed by the rod to pass entirely by a given point in the field of vision. We assume that P is always less than T; if the time of passage were longer than the oscillation time, the stroboscopic images would overlap and thus become very poorly defined.

If we now consider the **discontinuous variation** of fig. 8, we must distinguish between the four cases *I*, *IIa*, *IIb* and *III*, according as P is greater or smaller than D and $T—D$.

In each of the four cases we must now discover what is the greatest and what is the smallest brightness occurring in the field of vision. The brightness observed at a given point consists of the brightness of the background which amounts on an average to $B_0\,(1—ab)$ (for the meaning of a and b see fig. 8), plus an extra contribution caused by the passage of the lighter-coloured rod. To calculate these contributions we must remember that the time of passage is very short (for instance 0.005 sec) and that for such short times the eye, like the photographic plate, is capable of adding its impressions. From this it follows that we must calculate the extra contribution at a given point by integrating the brightnesses during the whole time of passage and dividing the result by the time M, which represents the longest time over which the eye still adds all impressions. This time M is of the order 1/20 sec. In this way for every point in the field of vision the observed brightness at each moment can be determined and then from all these brightnesses the smallest and the largest can be found and the most favourable position of the time of passage with respect to the dark period can be chosen.

In this way we calculate for the four cases the differences between the highest and the lowest brightnesses in the field of vision which are included in *table II*.

If we divide these brightness differences by the highest brightness occurring in the field of vision, which is also given in table II, we obtain finally the required contrast C in the field of vision, which is given in the last column of table II.

These four expressions for the contrast values enable us to survey the whole problem. We choose for the purpose first approximately those values for the different quantities which occurred in the experiment mentioned at the beginning, namely: $T = 0.01$ sec, $B = 10\,B_0$, $M = 0.05$ sec, $d = 0.5$ cm, $v = 100$ c/s and $P = 0.005$ sec, while we assume that $b = {}^1/_2$, $a = 0.3$ and therefore $D = 0.003$ sec.

Table II

	Brightness difference	Greatest brightness	Contrast C
I	$(B-B_0)\,\frac{(T-P)b}{M}$	$B_0(1-ab)+(B-B_0)\,\frac{P-b(P+D-T)}{M}$	$b(T-P) : \left[\frac{B}{B-B_0}\,(1-ab)\,M+P\,(1-b)+(T-D)b\right]$
IIa	$(B-B_0)\,\frac{Db}{M}$	$B_0(1-ab)+(B-B_0)\,\frac{P}{M}$	$bD : \left[\frac{B}{B-B_0}\,(1-ab)\,M+P\right]$
IIb	$(B-B_0)\,\frac{(T-D)b}{M}$	$B_0(1-ab)+(B-B_0)\,\frac{P-b(P+D-T)}{M}$	$b(T-D) : \left[\frac{B}{B-B_0}\,(1-ab)\,M+P\,(1-b)+(T-D)b\right]$
III	$(B-B_0)\,\frac{Pb}{M}$	$B_0(1-ab)+(B-B_0)\,\frac{P}{M}$	$bP : \left[\frac{B}{B-B_0}\,(1-ab)\,M+P\right]$

We then have case *IIa* and calculate $C = 0.154$. We then vary successively the quantities B/B_0, v, T, d, b and a and with the help of the formulae derived we determine the variation of the contrast. The result is reproduced in *fig. 9.* On each curve a point indicates the value from which we began in our example. Various conclusions may be drawn from these curves. Thus the first curve shows that the contrast can be appreciably increased by making B/B_0 larger. This can be achieved by choosing a glossier object. From the second and third curves we see that v and d were fairly well chosen; by increasing the velocity slightly or making the width of the rod slightly less, the contrast could have been increased somewhat. Furthermore it may be seen that the contrasts occurring in this experiment increase appreciably when T, b or a is increased.

From the relations found we may also deduce directly how, with T, b and a given, the other quantities must be chosen to obtain the effects as clearly as possible. In the first place B/B_0 must be made as large as possible; the rod must be as bright as possible against a background as dark as possible. Furthermore the time of passage P must be so chosen that we work in region *III*. This means that d/v must be chosen smaller than the smaller of the values aT and $(1-a)T$. Under these favourable conditions the maximum possible value of the contrast can be obtained, namely $C = b$, *i.e.* the same contrast which is seen to occur directly in fig. 8.

As a second example we consider a sinusoidal variation of the light intensity, as indicated in fig. 8. The calculation is the same; the integration over the time is somewhat more complicated, but on the other hand we are not compelled here to investigate four cases. The following general formula is found for the contrast C:

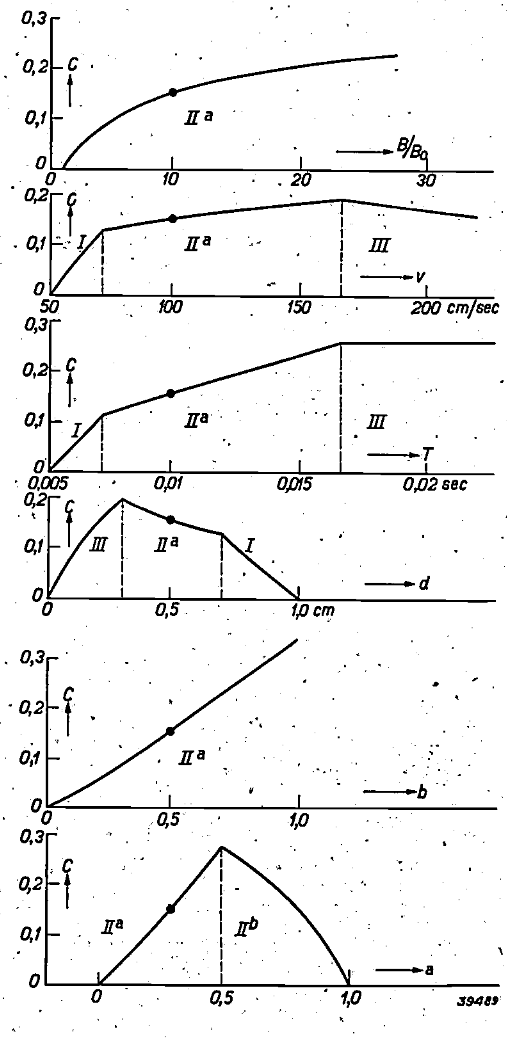


Fig. 9. Variation of the stroboscopic contrast C occurring in the field of vision upon illumination according to the discontinuous vibrational form of fig. 8, when one starts with an example (indicated by a point in the figures) and varies the following quantities successively:

B/B_0 = the ratio of the brightness of object and background.
v = the velocity with which the object moves.
T = the time of vibration of the light variation given in the first four sketches of fig. 8.
d = the width of the moving object measured in the direction of motion.
a and b = the quantities indicated in fig. 8.

$$C=\frac{2}{\pi}a_1\,T\sin\frac{\pi P}{T}:\left[M\frac{B_0}{B-B_0}+P+a_1\frac{T}{\pi}\sin\frac{\pi P}{T}\right]\cdot\ (2)$$

Here also we begin with a similar example, namely $T = 0.01$ sec, $a_1 = 0.5$, $B = 10\ B_0$, $M = 0.05$ sec, $d = 0.5$ cm, $v = 100$ cm/sec and $P = 0.005$ sec, and we again vary the different quantities. *Fig. 10* shows the results. Here also it may be seen that the contrast could have been increased by choosing B/B_0 larger. The quantities v and d are found to cause practically their maximum contrast value, while the contrast still increases appreciably when T and a_1 are increased.

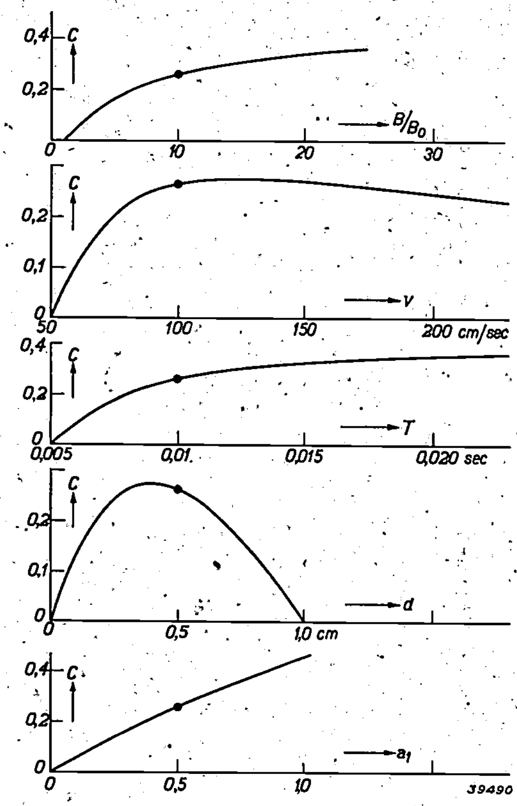


Fig. 10. Same as fig. 9, but for the continuous vibration form according to the last sketch of fig. 8.

If the quantities T and a_1 are given, the maximum possible contrast is again found by choosing B/B_0 as great as possible and $P = d/v$ as small as possible. The maximum value of C is then found to be $C = 2a_1/(1+a_1)$, *i.e.* again the same contrast as is read off directly from fig. 8.

Now that the contrasts are calculated for the different vibrational forms of fig. 8, we may ascertain to what extent the rule which is valid for the flickering of stationary objects may be applied: the non-sinusoidal light variation is resolved into Fourier components and only the fundamental tone is taken into account.

Upon the resolution of the discontinuous vibrational form drawn in fig. 8 we find:

$$a_1 = \frac{2b\sin\pi a}{\pi(1-ab)}\,;\qquad (3)$$

If the rule mentioned is correct, then by filling in this value in (2), a formula for C would be obtained which would give the same results as table II. From the nature of the formula we can see immediately that this can never be exactly the case. In order to obtain some idea of the magnitude of the errors which are made by thus neglecting the overtones, we calculate the contrast for several values of a and b in two ways, namely 1) neglecting the overtones and 2) with the help of table II. For the quantities T, B/B_0, M, d and v we assume the values which we have used in the two previous examples. *Table III* gives the results of these calculations. It may be seen that upon neglecting the overtones we always find too high values for C. When a is not too small, the deviations are not greater than 10 per cent; with very small values of a the deviations amount to about 25 per cent.

In the above we have been concerned with the calculation of the contrasts which occur. In order to judge whether a given value of C should be considered disturbing we must still determine by means of measurements what the threshold value for C is under these circumstances, in other words what contrast C can just barely be observed by the eye. At the same time these measurements will be a check on the correctness of the theory: when we vary different variables (a, b, v, d, a_1, etc.) until flicker can be observed, we must each time find about the same threshold value for C. Several preliminary measurements gave a threshold value of $C = 0.10$ at different illumination intensities, calculated on the assumption that $M = 0.05$ sec.

Until now we have only considered the case in

Table III

Values of the contrast C for different values a and b, when the overtones are neglected and with the help of the expressions given in table I, respectively.

b \ a	$\to 0$	0.3	0.7
$\to 0$	1.20 ab 0.95 ab	0.0311 b 0.0284 b	0.0311 b 0.0284 b
0.3	0.36 a 0.284 a	0.098 0.090	0.110 0.102[4]
0.7	0.84 a 0.66 a	0.242 0.224	0.351 0.326
1	1.20 a 0.95 a	0.364 0.338	0.682 0.642

which a rod crossed the field of vision once. The case is different when an object moves periodically with the same period as that with which the light flickers. Under these circumstances the stroboscopic images will usually be very clear. In many cases contrasts are then seen to occur which lie close to the maximum value $C = b$ and $C = 2a_1/(1+a_1)$.

THE SENSITIVITY OF AERIALS TO LOCAL INTERFERENCES

by P. CORNELIUS.

621.396.67

Although the use of a good outdoor aerial and earth connection is always to be recommended for good reception, their installation is sometimes impossible or undesirable for other reasons. Indoor aerials usually give an unfavourable ratio between signal and interference, and this objection is even more valid for a mains aerial. The latter type of aerial, however, has the advantage that the set can be connected at any wall contact, independent of an aerial connection. This advantage can also be realized with built-in and attached loop aerials.

In this article it is shown that a loop aerial is in most cases less sensitive to local interferences than a capacitative aerial with the same power of interception. For the sake of further diminishing the sensitivity to interferences, not only from local sources of interference but also from other transmitters, the directional effect of a loop aerial can often be used successfully.

Introduction

In ordinary radio receiving sets with mains connections, the use of a good outdoor aerial and a good earth connection is always recommended. Due to the great power of interception of this combination, weak transmitters can still be received with sufficient intensity. Moreover, the relation between signal and noise is favourable so that many stations can be received without noise, while the sensitivity to local interferences (from motors or caused by switches, and the like) is relatively low. The latter is due to the fact that the largest part of the outdoor aerial is at a relatively great distance from the lines of the electric installations of the house, along which the high-frequency interferences excited in the net are propagated.

The relation between signal and interference can be still further improved by the use of a shielded connection from the outdoor aerial to the set [1]).

Over against these advantages, however, the outdoor aerial has several practical disadvantages, namely:

1) In the installation of a good outdoor aerial and earth connection the help of an expert is usually necessary, which of course always involves some expense.
2) The installation of wires in a finished room and the assembly of the outdoor part of the aerial often encounters aesthetic objections.
3) The installation of an outdoor aerial is often impossible in large cities because of lack of space.
4) The receiving set must remain at the spot once chosen for the aerial connection, although it is often desirable to be able to move it from one room to another.

In connection with these objections attempts have often been made to construct efficient aerials which lack one or more of these disadvantages. We shall briefly discuss the commonest types of such aerials.

a) A possibility of eliminating the disadvantages indicated under 1) and 3) is offered by the installation of *a permanent indoor aerial with the water main used as earth connection.* An objection to this is the fact that a given set always has a considerably greater sensitivity to interferences with an indoor aerial than with a good outdoor aerial, due to the relatively short distance between aerial and light main. The sensitivity and the ratio between signal and noise will nevertheless usually be satisfactory.
b) In order to eliminate the disadvantages 2) and 4) *the earth connection may be omitted and the aerial may be given a length of only a few metres.* It may then simply be hung or laid on the floor near the spot where the set is to be placed. The

[1]) This of course causes a certain weakening of the effective signal, so that the sensitivity and the ratio between signal and noise are somewhat diminished, see Philips techn. Rev. 4, 320, 1939.

which a rod crossed the field of vision once. The case is different when an object moves periodically with the same period as that with which the light flickers. Under these circumstances the stroboscopic images will usually be very clear. In many cases contrasts are then seen to occur which lie close to the maximum value $C = b$ and $C = 2a_1/(1+a_1)$.

THE SENSITIVITY OF AERIALS TO LOCAL INTERFERENCES

by P. CORNELIUS. 621.396.67

Although the use of a good outdoor aerial and earth connection is always to be recommended for good reception, their installation is sometimes impossible or undesirable for other reasons. Indoor aerials usually give an unfavourable ratio between signal and interference, and this objection is even more valid for a mains aerial. The latter type of aerial, however, has the advantage that the set can be connected at any wall contact, independent of an aerial connection. This advantage can also be realized with built-in and attached loop aerials.

In this article it is shown that a loop aerial is in most cases less sensitive to local interferences than a capacitative aerial with the same power of interception. For the sake of further diminishing the sensitivity to interferences, not only from local sources of interference but also from other transmitters, the directional effect of a loop aerial can often be used successfully.

Introduction

In ordinary radio receiving sets with mains connections, the use of a good outdoor aerial and a good earth connection is always recommended. Due to the great power of interception of this combination, weak transmitters can still be received with sufficient intensity. Moreover, the relation between signal and noise is favourable so that many stations can be received without noise, while the sensitivity to local interferences (from motors or caused by switches, and the like) is relatively low. The latter is due to the fact that the largest part of the outdoor aerial is at a relatively great distance from the lines of the electric installations of the house, along which the high-frequency interferences excited in the net are propagated.

The relation between signal and interference can be still further improved by the use of a shielded connection from the outdoor aerial to the set [1]).

Over against these advantages, however, the outdoor aerial has several practical disadvantages, namely:

1) In the installation of a good outdoor aerial and earth connection the help of an expert is usually necessary, which of course always involves some expense.
2) The installation of wires in a finished room and the assembly of the outdoor part of the aerial often encounters aesthetic objections.
3) The installation of an outdoor aerial is often impossible in large cities because of lack of space.
4) The receiving set must remain at the spot once chosen for the aerial connection, although it is often desirable to be able to move it from one room to another.

In connection with these objections attempts have often been made to construct efficient aerials which lack one or more of these disadvantages. We shall briefly discuss the commonest types of such aerials.

a) A possibility of eliminating the disadvantages indicated under 1) and 3) is offered by the installation of *a permanent indoor aerial with the water main used as earth connection.* An objection to this is the fact that a given set always has a considerably greater sensitivity to interferences with an indoor aerial than with a good outdoor aerial, due to the relatively short distance between aerial and light main. The sensitivity and the ratio between signal and noise will nevertheless usually be satisfactory.
b) In order to eliminate the disadvantages 2) and 4) *the earth connection may be omitted and the aerial may be given a length of only a few metres.* It may then simply be hung or laid on the floor near the spot where the set is to be placed. The

1) This of course causes a certain weakening of the effective signal, so that the sensitivity and the ratio between signal and noise are somewhat diminished, see Philips techn. Rev. 4, 320, 1939.

capacity of the chassis toward the mains and the surroundings then acts as earth connection. Such an aerial is usually unsatisfactory; the sensitivity and the ratio of signal to noise become poor and the level of the local interference very high.

c) An ideal solution of the problems 1) to 4) is formed by *the mains aerial without earth connection.*

In this case, however, not only is the sensitivity usually small, but the level of interference is unusually high, since the aerial connection is directly connected for high-voltage to one of the mains lines. The mains aerial indeed usually gives quite unsatisfactory results.

The loop aerial

The loop aerial also forms a solution of the problems mentioned. Compared with the other forms of movable aerials mentioned above, the loop aerial is found to offer advantages with respect not only to the sensitivity but also to the level of interference.

In connection with the sensitivity the following must be noted.

In the development of receiving sets which are intended for use with an aerial, the electrical data of the aerials of the users of the sets are never known. In order to avoid unallowable detuning of the first tuned circuit upon connection to different aerials, the aerial must be relatively loosely coupled with this circuit. Because of this, however, only a part of the energy which the aerial can deliver is given off to the first circuit. On the other hand the electrical data of a loop aerial permanently built in or attached to the set are known, so that it may be coupled very tightly with the first circuit without danger of detuning by capacity variation. Because of this a well constructed loop aerial usually gives a greater sensitivity and a better ratio of signal to noise than a small movable "ordinary" aerial. Moreover, it is found that a loop aerial often picks up much less interference than another aerial which exhibits the same reception strength for a distant station. In order to be able to explain this remarkable phenomenon, we shall in the following consider several properties of the electromagnetic field of transmitters and sources of interference, as well as several properties of receiving aerials.

The electromagnetic field of sources of electrical and magnetic radiation

An alternating electromagnetic field can be excited by different types of sources of radiation, among which electrical and magnetic sources of radiation may be distinguished as limiting cases. The former sources of radiation in the first instance excite alternating electric fields by supplying an A.C. voltage to an open condenser (aerial), while the latter excite alternating magnetic fields in the first instance with the help of an alternating current, which flows through a coil with a large surface. Due to the variability of these fields a magnetic field occurs also in the case of the electrical source; while conversely, the magnetic source will also excite an electric field. These fields, however, only become of importance at some distance from the source of radiation, as we shall see below. The commonest type of transmitter is the electrical type, which in an idealized form may be represented as a dipole with the capacity concentrated at the ends (*fig. 1*). With the help of Maxwell's equations

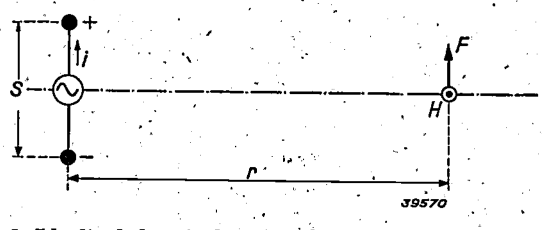


Fig. 1. Idealized electrical source of radiation, represented as a dipole with capacity concentrated at its ends. F and H represent the electric and magnetic field strengths, r is the distance from the transmitter to the point considered. H is directed forwards and thus indicated by a circle surrounding a point.

Hertz has calculated the strength of the electric field F and the strength of the magnetic field H in the electromagnetic field of such a dipole. Upon the use of Giorgi units the equations for F and H at a point lying in the plane bisecting the dipole (see fig. 1) are as follows:

$$F = \frac{1}{4\pi\varepsilon_0}\left(\frac{p}{r^3} + \frac{\dot{p}}{cr^2} + \frac{\ddot{p}}{c^2r}\right) \text{ (volts/metre)}, \quad (1)$$

$$H = \frac{c}{4\pi}\left(\frac{\dot{p}}{cr^2} + \frac{\ddot{p}}{c^2r}\right) \text{ (amperes/metre)}, \quad (2)$$

where:

p is the electrical moment of the dipole (in coulomb $\times$ metre), which is defined as the product of the magnitude of the two electric charges and their separation. This electrical moment is equal to the couple in watt seconds (1 watt sec $=$ 10^5 dyne metre) which the dipole would experience if it were placed in an electric field with a strength of 1 V/m, the lines of force of which are perpendicular to the axis of the dipole;

$\dot{p}$ is the first derivative of p with respect to the time in seconds;

$\ddot{p}$ is the second derivative of p with respect to the time;

$c = 3 \times 10^8$ m/sec, the velocity of light;

r is the distance from the point considered to the dipole in metres:

$\varepsilon_0 = \frac{10^7}{4\pi c^2} = \frac{1}{36\pi} \cdot 10^{-9} \left(\frac{\text{A sec}}{\text{Vm}}\right)$, the so-called absolute dielectric constant, *i.e.* the dielectric constant for vacuum.

If we assume that between the ends of the dipole which are separated by a distance s a sinusoidal alternating current ($i_0 e^{j\omega t}$) with the angular frequency ω is flowing, then

$$p = \frac{i}{j\omega} s \text{ (Asec m)},$$
$$\dot{p} = i\,s \text{ (A m)},$$
$$\ddot{p} = j\,\omega\,i\,s \text{ (Am/sec)}.$$

The formulae given are valid only when $s \ll r$ and $s \ll \lambda$, where λ is the wave length.

Formulae (1) and (2) may also be used for transmitters in which the yield is generated between an aerial and earth. The dot-dash line in fig. 1 may then be considered as the ideal conducting earth's surface perpendicular to the axis of the dipole. The lower half of the figure is then no longer considered, while the fields in the upper half experience no change.

The magnetic source of radiation can be represented in an idealized form by a coil consisting of n windings in which a current flows (*fig. 2*).

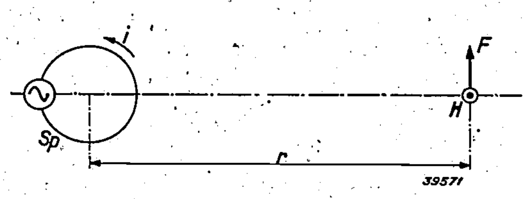


Fig. 2. Idealized magnetic transmitter, represented by a coil of n windings with a surface O; the axis of the coil is perpendicular to the plane of the drawing; F and H are the strengths of the electric and magnetic fields in the field of the travelling wave. r is the distance between the transmitter and the point considered. H is directed forward and thus indicated by a circle around a point.

The formulae of the electrical and the magnetic field strength at a point lying in the plane of the coil are given by the following equations:

$$H = \frac{1}{4\pi\mu_0}\left(\frac{m}{r^3} + \frac{\dot{m}}{cr^2} + \frac{\ddot{m}}{c^2 r}\right) \text{ (A/m)} \quad . \; . \quad (3)$$

$$F = \frac{c}{4\pi}\left(\frac{\dot{m}}{cr^2} + \frac{\ddot{m}}{c^2 r}\right) \text{ (V/m)} \quad . \; . \; . \; . \; . \quad (4)$$

where:

m is the magnetic moment of the coil, which is defined as the product of the surface of the coil (in m²), the number of windings and the current (in A). This magnetic moment is equal to the couple in watt sec which the coil would experience if it were placed in a magnetic field of 1 A/m whose lines of force lie in the plane of the coil.

$\dot{m}$ is the first derivative of m with respect to the time in sec.

$\ddot{m}$ is the second derivative of m with respect to the time.

$\mu_0 = 4\pi \cdot 10^{-7} \left(\frac{\text{V sec}}{\text{Am}}\right)$, the so-called absolute magnetic permeability.

For a sinusoidal alternating current i

$$m = \mu_0\, n\, i\, O \text{ (Vsec m)},$$
$$\dot{m} = \mu_0\, n\, i\, O\, j\omega \text{ (Vm)},$$
$$\ddot{m} = \mu_0\, n\, i\, O\, j^2\omega^2 \text{ (Vm/sec)}.$$

Here also the dimensions of the source of radiation must be small compared with r and λ.

For the reader who is not accustomed to the use of the absolute dielectric constant and the absolute permeability, the following remarks may be useful.

The capacity of a plane condenser, in which the dimensions of the condenser plates (O) are large compared with their distance apart d, is

$$C = \varepsilon_0 \frac{O}{d} \left(\frac{\text{A sec}}{\text{V}} = \text{farads}\right),$$

where O/d is expressed in metres.

If a direct current i is sent through a long cylindrical coil with a cross section O and a length l over which n windings are uniformly distributed, the magnetic field in the coil is ni/l (A/m). This field is directed according to the axis of the coil in such a way that the direction of the current in the coil and the direction of the field together give the motion of a right-hand screw. It is hereby assumed that the dimensions of the cross section of the coil are small compared with the length, *i.e.* that, except at the ends, the field in the coil is practically homogeneous. The self-induction of this coil is

$$L = \mu_0 n^2 \frac{O}{l} \left(\frac{\text{Vsec}}{\text{A}} = \text{henrys}\right),$$

where O/l is again expressed in metres.

Comparison of the electrical and magnetic sources of radiation

If we consider a point which lies at such a great distance from an electrical source of radiation that

$$\left|\frac{\ddot{p}}{c^2 r}\right| \gg \left|\frac{\dot{p}}{cr^2}\right| \gg \left|\frac{p}{r^3}\right| \quad i.e. \quad r \gg \frac{\lambda}{2\pi},$$

then, according to (1) and (2)

$$F = \frac{1}{4\pi\varepsilon_0}\frac{\ddot{p}}{c^2 r} \text{ (V/m)}, \quad . \; . \; . \; . \quad (5)$$

$$H = \frac{c}{4\pi}\frac{\ddot{p}}{c^2 r} \text{ (A/m)}. \quad . \; . \; . \; . \; . \quad (6)$$

The ratio between F and H is therefore

$$\frac{F}{H} = \frac{1}{\varepsilon_0 c} = 4\pi c \cdot 10^{-7} \text{ (V/A)} \quad . \quad . \quad (7)$$

If we now consider a point at such a distance from a magnetic source of radiation that

$$\left|\frac{\ddot{m}}{c^2 r}\right| \gg \left|\frac{\dot{m}}{cr^2}\right| \gg \left|\frac{m}{r^3}\right|,$$

thus $r \gg \lambda/2\,\pi$ again, then according to (3) and (4)

$$H = \frac{1}{4\pi\mu_0} \frac{\ddot{m}}{c^2 r} \text{ (A/m)}, \quad . \quad . \quad . \quad . \quad (8)$$

$$F = \frac{c}{4\pi} \frac{\ddot{m}}{c^2 r} \text{ (V/m)}. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

and therefore once more

$$\frac{F}{H} = \mu_0 c = 4\pi c \cdot 10^{-7} \text{ (V/A)}. \quad (10)$$

From (7) and (10) it follows that at a sufficiently great distance from an electrical and a magnetic transmitter there is a constant ratio between the electrical and the magnetic field strength which is the same in both cases. The electromagnetic field at such a distance from the transmitter is called the field of the travelling wave. In this field F and H are in phase and are perpendicular to each other and to the direction of propagation. The latter corresponds to the displacement of a right-hand screw which is turned from F toward H. When the observer is so far away from the transmitter that he is only concerned with this field, the electrical and the magnetic transmitters for the state drawn in figs. 1 and 2 could be interchanged without producing any effect on the field at the position of the observer. If i has the same value in both cases the following would have to be true for a coil with one winding:

$$O = \frac{c}{\omega} s = \frac{\lambda}{2\pi} \cdot s$$

while the phase in the transmitters would have to be so adjusted that the upper half of the dipole has its largest positive voltage at the same moment as the current in the coil, flowing to the left, reaches its maximum value.

We shall now consider a point lying much closer to the transmitter, so close in fact that the following holds in the case of the electrical transmitter:

$$\left|\frac{\ddot{p}}{c^2 r}\right| \ll \left|\frac{\dot{p}}{cr^2}\right| \ll \left|\frac{p}{r^3}\right|$$

and in the case of the magnetic transmitter:

$$\left|\frac{\ddot{m}}{c^2 r}\right| \ll \left|\frac{\dot{m}}{cr^2}\right| \ll \left|\frac{m}{r^3}\right|,$$

which in both cases amounts to $r \ll \lambda/2\pi$. In this case for the electrical transmitter

$$F = \frac{1}{4\pi\varepsilon_0} \frac{p}{r^3} \text{ (V/m)}. \quad . \quad . \quad . \quad . \quad (11)$$

$$H = \frac{\dot{p}}{4\pi r^2} \text{ (A/m)}, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

so that the ratio of F to H becomes

$$\left|\frac{F}{H}\right| = \left|\frac{p}{\varepsilon_0 r \dot{p}}\right| = \frac{1}{\varepsilon_0 c} \cdot \frac{\lambda}{2\pi r} \text{ (V/A)} \quad . \quad . \quad (13)$$

From (7) and (13) it follows that in the vicinity of the electrical transmitter the ratio of electrical to magnetic field is $\dfrac{\lambda}{2\pi r}$ times as great as it is at a large distance in the region of the travelling wave.

In the neighbourhood of a magnetic transmitter

$$H = \frac{1}{4\pi\mu_0} \frac{m}{r^3} \text{ (A/m)}, \quad . \quad . \quad . \quad . \quad (14)$$

$$F = \frac{\dot{m}}{4\pi r^2} \text{ (V/m)}, \quad . \quad . \quad . \quad . \quad . \quad (15)$$

so that the ratio of F to H becomes

$$\left|\frac{F}{H}\right| = \left|\frac{\mu_0 r \dot{m}}{m}\right| = \mu_0 c \frac{2\pi r}{\lambda} \text{ (V/A)} \quad . \quad . \quad (16)$$

From (10) and (16) it follows that in the neighbourhood of a magnetic transmitter the ratio of electrical to magnetic field is $\dfrac{\lambda}{2\pi r}$ times as small as at a great distance in the region of the travelling wave.

From the above the well-known fact becomes evident, that at a sufficiently small distance from an electrical source of radiation practically only an electrical field occurs, and in the neighbourhood of a magnetic source of radiation practically only a magnetic field. Actually these fields are described by the laws of Coulomb and Biot-Savart.

Capacitive and inductive receiving aerials

Just as in the case of the transmitters the electrical and the magnetic source of radiation can be distinguished as limiting cases, in the same way in reception a distinction can be made between capacitive and inductive aerials.

An ordinary capacitive aerial may, as in fig. 1, be represented by one half of a dipole perpendicular to the earth's surface with a capacity C toward earth concentrated at one end (*fig. 3*). If this

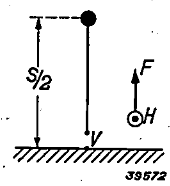


Fig. 3. Capacitive aerial represented diagrammatically as a half dipole with capacity concentrated at the top. $s/2$ is the distance from the top to earth, F and H are the electric and magnetic fields at a point.

aerial is situated in a field of radiation with an electrical field strength F, we obtain at the open lower end an A.C. voltage

$$V = \int_0^{s/2} F\mathrm{d}s = F\,\frac{s}{2}\ (\mathrm{V}); \quad . \quad . \quad (17)$$

V is thus independent of ω. The short-circuit current of the aerial amounts to:

$$I = V \cdot \omega \cdot C\ (\mathrm{A}) \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The magnitude of the magnetic field H indicated in fig. 3 has no effect on V or I.

This aerial thus forms an indicator for the intensity of an alternating electrical field in the direction of the dipole.

An inductive receiving aerial (loop aerial) is represented diagrammatically in *fig. 4* by a coil consisting of one winding. The self-induction of the coil is L_1, the surface O and the axis of the coil is parallel to H.

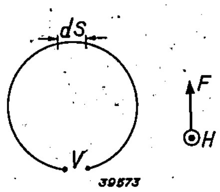


Fig. 4. Idealized representation of an inductive receiving aerial (loop aerial). F and H are the electric and magnetic fields at a point.

In an alternating magnetic field the magnetic lines of force are surrounded by closed electrical lines of force of which the line integral of the field strength along a closed curve, for instance, along the coil in question, is given by Maxwell's second law:

$$V = \int F\mathrm{d}s = \mu_0\, O\dot{H}\ (\mathrm{V}). \quad . \quad . \quad . \quad . \quad (19)$$

The magnitude of the electrical field drawn in fig. 4 has no effect on this. The loop aerial thus forms an indicator of the intensity of an alternating magnetic field in the direction of the normal to the plane of the loop.

If we assume a sinusoidally alternating magnetic field we obtain between the open ends of the coil an A.C. voltage:

$$V = \mu_0 \cdot O \cdot \omega H\ (\mathrm{V}), \quad . \quad . \quad . \quad . \quad (20)$$

while the short-circuit current of the coil is

$$I = \frac{V}{\omega L_1} = \mu_0 \cdot \frac{O}{L_1} \cdot H\ (\mathrm{A}) \quad . \quad . \quad . \quad (21)$$

The latter is thus independent of ω.

If, retaining its form, the coil is made not of one but of n windings, the voltage becomes larger by a factor n, while the short-circuit current becomes a factor n smaller, because the self-induction has increased by n^2 ($L_n = n^2 L_1$). If we further assume, that we are concerned with the field of radiation of a travelling wave, we can express H and F with the help of equation (7) and thereby obtain for the open voltage V and the short-circuit current I:

$$V = \frac{\omega}{c} \cdot n \cdot O \cdot F\ (\mathrm{V}),$$

$$I = \frac{1}{c} \cdot \frac{nO}{L_n} \cdot F\ (\mathrm{A}).$$

It must still be noted that the formulae given are only valid when the dimensions of the aerials are small compared with the wave length considered.

The field of a local source of interference

A local source of interference, for instance a vacuum cleaner with a sparking motor, excites irregular high-frequency oscillations. The reception of a radio set in the vicinity will be disturbed by components of these oscillations which correspond to the wave length selected. Since the source of interference is usually at a short distance from the set affected, the ratio F/H may differ very much from the value of the ratio for distant transmitters. It is now important to know whether F/H is as a rule smaller or larger for a source of interference than in the field of a travelling wave. In the former case, in order to obtain reception free of interference, the capacitive aerial will be preferable, in the latter case the loop aerial. In this connection therefore it is necessary to discover whether a source of interference behaves mainly as an electrical or as a magnetic source of radiation. This is found to depend upon the course of the connecting wires.

It is practically exclusively the so-called "asymetrical" components of the interference currents excited which are disturbing, *i.e.* a current which, *via* the capacity of the source of interference toward earth and *via* the connecting wires coming out of the ground and partially surrounded by metal tubes, flows back to the source of interference (*fig.* 5)[2]).

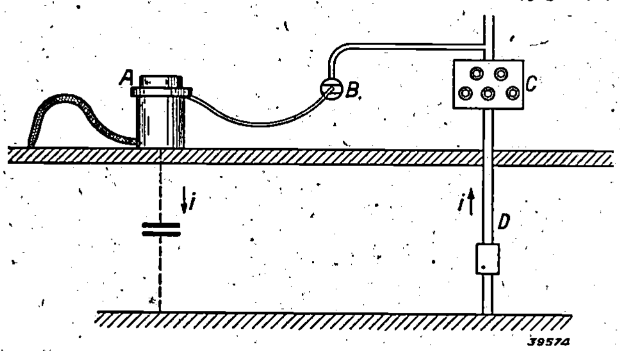


Fig. 5. Course of the asymmetrical component of the interference current from a sparking vacuum cleaner on the first floor. *A* vacuum cleaner, *B* wall contact, *C* switch board, *D* main line of the mains.

Represented in an idealized form the four cases of different positions of the wires shown in *fig.* 6 may occur.

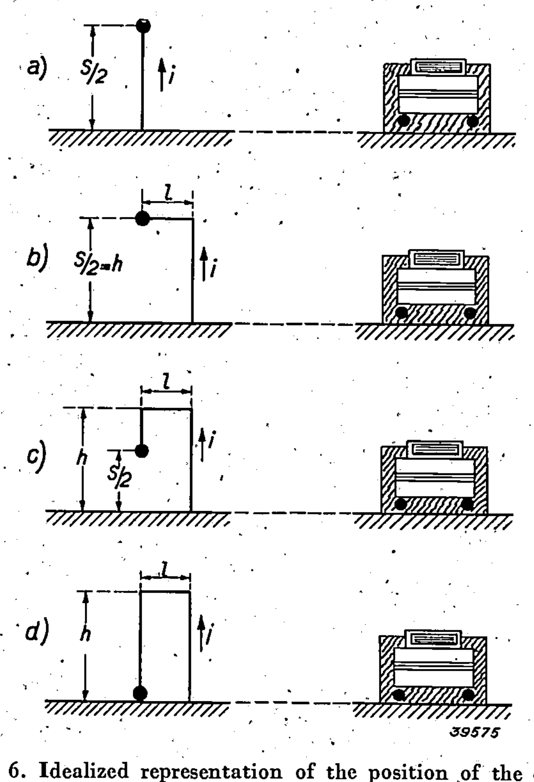


Fig. 6. Idealized representation of the position of the connection lines to a source of interference. The source of interference behaves as an electrical transmitter in case *a* and as a magnetic transmitter in case *d*. In cases *b* and *c* it behaves as a combination of electrical and magnetic transmitter. The set sketched represents the position of the receiving aerial (distance *r* from the source of interference).

In case *a* the source of interference behaves as an electrical transmitter, while in case *d* it will behave as a magnetic transmitter. In cases *b* and *c*, on the other hand, which are the commonest cases, the source of interference will behave partially as a magnetic and partially as an electrical transmitter. The relation between these two, and thus also the relation between the electrical and the magnetic field of the interference caused, depends upon the relation between l, h and $s/2$.

We shall now make a comparison between the interference which is experienced upon the use of a receiving set with a loop aerial and a set with a capacitive aerial which are situated at the point indicated in fig. 6 with respect to a source of interference. For this purpose we assume that the sets are so arranged that when situated in the travelling wave of a transmitter, which is in the same direction as the source of interference with respect to the set, they give the same grid A.C. voltage at the first valve. If the receiving sets are so close to the source of interference that $r \ll \frac{\lambda}{2\pi}$, then in the case of fig. 6*a*, the relation between signal and interference will be better for the loop than for the capacitive aerial, because the ratio of the electrical field strength F, to which the capacitive aerial reacts, to the magnetic field strength to which the loop aerial reacts is, according to (13), $\lambda/2\pi r$ times as great as in the travelling wave of the transmitter to be received.

For case *d* the capacitive aerial would be better than the loop; this case will, however, seldom occur in practice. For cases *b* and *c*, where we are concerned with a combination of an electrical and a magnetic transmitter, we shall calculate what the dimensions of l and h must be in relation to $s/2$ to give F and H at the distance r the same ratio as in the field of a travelling wave. In this case the

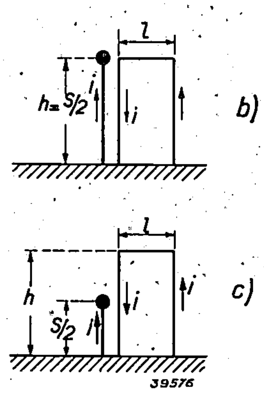


Fig. 7. Substitution diagrams for the figures 6 *b* and *c*. The open loop is closed; the final capacity is situated at practically the same point in space, but now takes the form of the final capacity of half a dipole, the vertical conductor of which is situated in the immediate neighbourhood of the part of the loop which has been added. The current in this dipole is equal in strength and opposite in direction to that in the part of the loop added.

---

[2]) See Philips techn. Rev. 3, 235, 1938.

sensitivity to interferences will be the same for both sets.

The cases b and c from fig. 6 are equivalent to the substitution diagrams drawn in *fig. 7b* and *c*. The open loop is here replaced by a closed loop (*i.e.* a magnetic source of radiation) and a half of a dipole in the immediate vicinity of the added section of the loop. The current in this dipole is equal in magnitude and opposite in direction to that in the added section referred to.

In the neighbourhood of the source of radiation the electrical field F_s is determined practically exclusively by the electrical moment of the combination. This field strength is given by (11) in which $p = i\,s/\omega$. The magnetic field, which is determined practically exclusively by the magnetic part of the combination, is given in the vicinity by (14), where, taking into account the reflection of a perfectly conducting earth $m = 2\,\mu_0\,i\,l\,h$.

Since we wish to set the ratio of F_s and H_s equal to that in a travelling wave, according to equation (7) the following is valid:

$$F_s = \frac{1}{\varepsilon_0 c}\,H_s.$$

By the substitution of F_s and H_s from (11) and (14) in this equation we find the following upon the use of the above values for p and m:

$$\frac{1}{4\pi\varepsilon_0}\cdot\frac{i\,s}{\omega r^3} = \frac{1}{\varepsilon_0 c}\cdot\frac{1}{4\pi\mu_0}\cdot\frac{\mu_0\,i\,l\cdot 2h}{r^3}$$

and thus

$$l = \frac{c}{\omega}\cdot\frac{s}{2h} = \frac{\lambda}{2\pi}\,\frac{s}{2h}\,.$$

The horizontal length of the connection line would, therefore, in case b where $h = s/2$, have to be equal to $\lambda/2\pi$ before the loop aerial exhibits a disadvantage compared with the capacitive aerial with respect to sensitivity to interference. If for case c we assume that $h = s$, this would be the case at a length l of $\lambda/4\pi$. For the intermediate and long-wave regions a length of twenty to forty metres would be necessary, which of course seldom occurs.

In the above it was assumed that the loop aerial is situated in the most unfavourable position with respect to the source of interference, namely that the desired transmitter and the source of interference are in the same direction from the receiving set. When this is not the case the effect of local interference can often be considerably decreased by making use of the well-known directional effect of the loop aerial.

In practice it is found in most cases that for the intermediate and long-wave regions the sensitivity for interferences of a well constructed loop aerial is appreciably smaller than that of a capacitive indoor aerial.

HARD GLASS X-RAY TUBES IN OIL

by J. H. van der TUUK. 621.386.1

The use of hard glass has made it possible considerably to decrease the dimensions of all-glass X-ray tubes. When the tube is placed in an earthed metal jacket in order to protect the user against high voltage, the space between tube and jacket may be filled with oil which promotes the voltage security and the heat dissipation. By a suitable construction of the anode it is possible to adapt the tube to very high continuous loads. Various X-ray tubes developed on these principles for medical diagnosis, therapy and the testing of materials are described in this article.

The X-ray tubes in general use ten or twenty years ago were actually not "tubes" but rather "bulbs", considering the form of the glass container in which the electrodes were placed (*fig. 1*). This

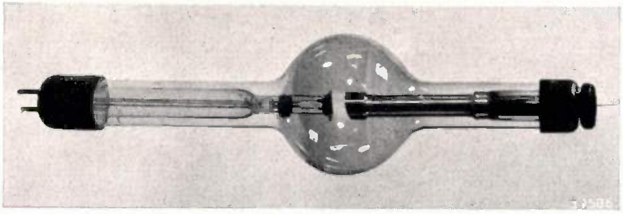


Fig. 1. An X-ray tube such as was used around 1920. Because of the bulbous shape the glass container suffers less from the bombardment by secondary electrons.

bulb form sprang from the desire to keep the glass wall at some distance from the electrode on which the X-radiation was excited by electron bombardment. If the distance was too small the glass wall exhibited very intense fluorescence which was found to be unfavourable for the life of the instrument: the bombardment of the wall by secondary electrons, which was the cause of the fluorescence caused at the same time a gradual destruction of the glass by heating and electrolytic attack, and freed gases from the glass which gradually spoiled the vacuum in the tube and finally usually caused breakdown through the glass.

Increasing the distance between electrode and wall in an attempt to decrease this danger, was of itself not a very fortunate solution. The disadvantages became particularly evident when it was understood that it was necessary in practice to protect the user absolutely from the X-radiation outside the effective beam and from the high voltage. The bulbous form of the X-ray tubes made it difficult to introduce the necessary lead jackets and earthed metal coverings since the resulting constructions were much too large and expensive.

A fundamental improvement in the situation was first presented by the construction of the so-called metal tubes [1]). In such tubes the discharge space from which the secondary electrons originate is surrounded by a chrome-iron cylinder to which glass sleeves are welded to serve as insulation for the electrodes. Since the metal cylinder is extremely insensitive to the electron bombardment, the diameter of the cylinder could be small.

Later it was also found possible to decrease the dimensions of the all-glass tubes considerably, by the use of **hard glass**. Hard glass is in general much more resistant to electron bombardment since it is not so strongly electrolysed as soft glass, and it has a much greater electrical resistance and resistance to breakdown. Due to the high resistance the glass, which is negatively charged by an electron bombardment, retains the charge and thereby repels further electrons. Moreover, hard glass has a much smaller linear coefficient of thermal expansion than ordinary soft glass, namely about $45 \times 10^{-7}/°C$ instead of about $90 \times 10^{-7}/°C$, so that it has less tendency to break upon local heating.

Just as in the construction of the tubes with a metal intermediate section it was a necessary condition that the chromeiron should have about the same coefficient of expansion as the ordinary kinds of glass used with it, in the construction of the hard-glass tubes it was also of great importance for making the relatively large anode leads to have at one's disposal alloys whose thermal expansion is as well adapted to that of the hard glass. An alloy of iron, nickel and cobalt satisfies this requirement over the whole temperature region from about 0 °C to beyond the softening interval of the glass.

When it is taken into account that the fundamental idea, with the "metal" as well as with the hard-glass tubes, is to make the tube wall which is struck by secondary electrons of a material which is more resistant to this bombardment, and that in both cases devices are also employed to limit the number of secondary electrons which can reach the tube wall, there no longer seems to be very much difference between the two types of tube. Nevertheless in the further development of the tubes, when

[1]) A. Bouwers, Fortschr. Röntgenstr. **32**, 41, 1924; Radiology **13**, 191, 1929.

it is a question of making a unit which is secure against high voltage and radiation, there are different constructive possibilities in the two cases. In the case of the former (metal) tubes the earthed jacket, usually filled with a gaseous or liquid insulation material, can be fastened to the chrome-iron intermediate section which is likewise earthed. This then serves as a natural support for the suspension of the tube, while the lead jacket which gives protection against the rays is placed directly over this. In the case of the hard-glass tubes the jacket must be more free of the tube. Usually the space between the tube and the earthed metal jacket is filled with a liquid insulation material, preferably transformer oil, which has favorable properties in connection with voltage security and heat dissipation.

Various types of "metal" tubes have already been discussed in this periodical [2]). We shall now also describe several types of hard glass X-ray tubes in oil. We shall hereby confine ourselves to the tubes with stationary anode.

Tube for medical diagnosis

Voltage security

In *fig. 2* a diagram is given of a cross section of a hard-glass X-ray tube manufactured by Philips for general purposes of diagnosis. *Fig. 3* is a photograph of such a tube. The small dimensions of the tube are immediately striking: the glass tube is about 13 cm long. Nevertheless, this short tube is suitable for the highest D.C. and A.C. voltages used in diagnosis, namely up to 100 kV_{max}, thanks in part to the high resistance to breakdown of the oil in which the tube is immersed when in use. This resistance to breakdown, which amounts to 200 kV/cm for good fresh transformer oil, can with proper use scarcely fall below 60 kV/cm even for long used oil, and it always remains about 15 kV/cm even along the path of the creeping discharge on the glass wall. A breakdown along the outside is thus impossible at voltages of 100 kV, and indeed the limit of voltage security of the tube is not determined by this, but by the phenomena taking place inside the tube. With D.C. voltage it is mainly a question of the bombardment of the tube wall by secondary electrons, with A.C. voltage there is the added danger of back-lash.

In order to stop the secondary electrons which are freed from the focus on the anode, the front of the anode (the tungsten "lozenge" on which the beam of primary electrons is focussed) is surrounded by a copper shielding cap *H* which has only two relatively

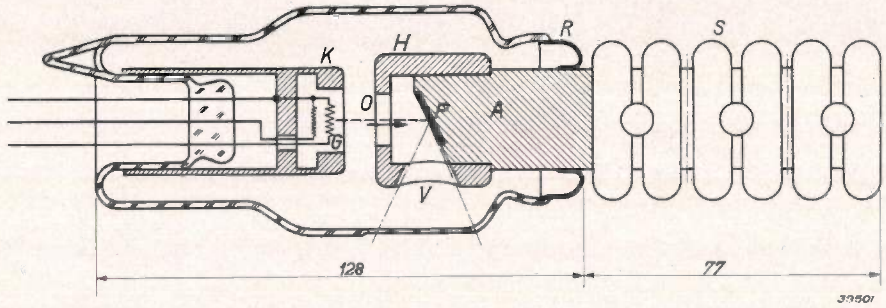


Fig. 2. Cross section of a hard-glass tube for diagnosis, somewhat simplified. The cathode *K* with filament *G* is fastened into the tube by means of a bridge, while the anode *A* is welded into the tube *via* a ring *R* of an iron-nickel-cobalt alloy. On the sloping front surface of the anode is the "lozenge" *P* of tungsten upon which the electrons from *G* are focussed. These electrons pass through the opening *O* of the anode cap *H* which captures the secondary electrons liberated from the focus. The X-rays leave the tube through the opening *V*. For cooling, the radiator *S* is fastened to the anode on the outside. Its dimensions are determined by the desired power for fluoroscopy (see below). With the radiator here drawn the tube weighs about 1.5 kg.

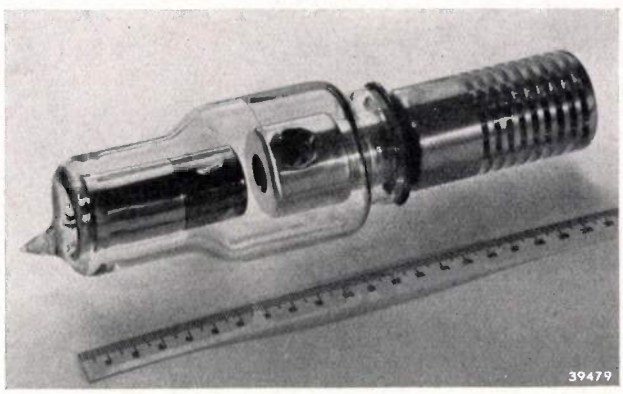


Fig. 3. Photograph of the X-ray tube whose cross section is shown in fig. 2. (Anode seal and radiator are slightly different here).

[2]) J. E. de Graaf and W. J. Oosterkamp, An X-ray tube for the analysis of crystal structure, Philips techn. Rev. 3, 259, 1938; J. H. van der Tuuk, A million volt X-ray tube, Philips techn. Rev. 4, 153, 1939; H. A. G. Hazeu, Small apparatus for medical X-ray diagnosis, Philips techn. Rev. 6, 225, 1941.

narrow openings: one at O to admit the primary electrons and one at V to permit the effective X-ray beam to leave the tube (the cap thus acts at the same time as a certain protection against X-rays in other directions, although its protection is not sufficient). The latter opening is not larger than corresponds to an apex angle of $2 \times 10°$ of the cone of the X-rays, since the front surface of the anode slopes 10° [3]), see fig. 2. Because of the small size of this opening it was found to be unnecessary to screen the secondary electrons passing through it from the walls by means of a metal foil; the effective X-rays therefore do not need to be attenuated by such a foil. At the same time there is the advantage that it is possible to look through the opening at the focus and thus detect any defects. The distance between the anode cap and the anode is 10 mm, so that, taking into account the presence of slight irregularities, the field strength on the electrodes at 100 kV still remains far below the limit of 10^7 V/cm, at which electrons can be liberated by "cold emission" from clean metal surfaces [4]). This source of stray electrons which might reach the tube wall is therefore also eliminated.

As to the danger of back-lash when working on A.C. voltage, this is also considerably diminished by the presence of the anode cap. The back-lash is based upon the possibility that in the negative phase, *i.e.* the half of the period in which the filament is positive and the lozenge is negative, electrons may be drawn out of the still hot focus and, accelerated by the tube voltage of for instance 100 kV, reach the delicate filament — in this case a tungsten wire of 200 μ diameter. This will happen for example when, due to an accidental heavy overloading in the positive phase, the focus has become temporarily too hot, either as a whole or locally. By this undesired electron bombardment the filament is now overheated in the negative phase, in the next positive phase the emission therefore becomes higher, and the focus therefore still more heavily loaded, etc. The immediate result is therefore an avalanche-like process of stronger and stronger overheating of lozenge and filament, so that in general the filament burns through and the tube becomes defective. A suitably constructed anode cap very much decreases the field strength at the focus and thus the chance of back-lash.

Another method of combatting back-lash consists in the excentric placing of the filament. Any electrons emitted by the focus in the negative phase then do not fly to the filament but strike other, less vulnerable parts of the cathode. In our case also advantage was taken of this circumstance, since the tube possesses a so-called double focus: at the cathode there are two filaments side by side (thus slightly excentric) from which a selection can be made and which give respectively a normal focus of 3.1×3.1 mm² or a very small focus of 0.3×0.3 mm² for special fluoroscopy or photography techniques.

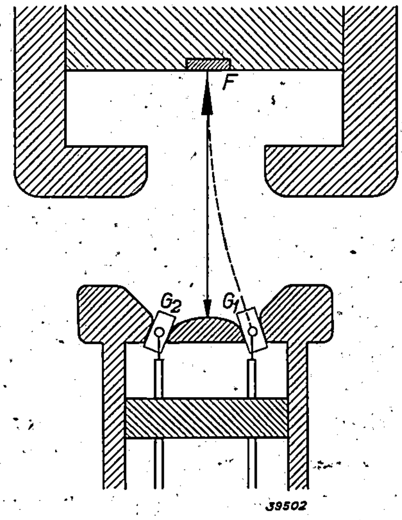


Fig. 4. The cathode of the tube described contains two spiral filaments G_1 and G_2 either of which can be switched on as desired in order to obtain foci of different sizes. Due to the centric position of the filaments there is less chance of back-lash, since any electrons drawn out of the hot focus F by the electric field do not strike the filament, but some other spot on the cathode. This does not alter the fact that with the reverse field the electrons from the filaments fly to the focus, because the trajectories of the electrons are in general not reversible due to inertia. (See J. H. van der Tuuk, Physica 10, 231, 1930).

Permissible loading for photography

In making X-ray photographs, definite standardized momentary loads on the tube are used, which loads depend upon the size of the focus employed. With a focus of 3.1×9.3 mm² true size (*i.e.* with a slope of the lozenge of 19° an apparent size of 3.1×3.1 mm²) these instantaneous loads are 6 kW for 1 sec and 4 kW for 5 sec when used with pulsating D.C. voltage. The lozenge is thus loaded with about 200 watt/mm² in exposures of 1 sec. By embedding the tungsten lozenge in a large mass of copper the temperature of the focus is limited to about 2 300 °C at such a load, a temperature which is still quite permissible in connection with evaporation.

The focus temperature becomes established so

[3]) This slope means that a linear focus is used with a length three times its width; see the first article referred to in footnote [2]).

[4]) See for example W. Ch. van Geel, Blocking-layer Rectifers, Philips techn. Rev. 4, 100, 1930.

quickly that it may be said to be determined by the momentary value of the load. Consequently upon the use of A.C. voltage the peak value of the current may not be larger than approximately the current with D.C. voltage. On the other hand the rapid establishment of the focus temperature has as a result that with a maximum temperature of about 2 300 °C in the positive period the focus is already sufficiently cooled to cause no back-lash in the negative period; the peak current with A.C. voltage need not therefore be smaller than the current with D.C. voltage. The permissible average specific loading with A.C. voltage is thus about half of that with D.C. voltage: the standardized momentary loads for the sizes of focus mentioned are here 3 kW for 1 sec and 2.2 kW for 5 sec.

During the exposure only little heat is given off to the surroundings by the anode, the anode must therefore have sufficient heat capacity so that it does not become warm as a whole with the brief high loads indicated. In the tube in question the anode weighs about 1 000 g including the radiator visible in fig. 2, the specific heat of copper is 0.094 cal/g °C, the heat capacity is thus 94 cal/°C. With a D.C. voltage load of 4 kW for 4 seconds the anode receives a total of 20 000 Wsec $\approx$ 5 000 cal. The temperature of the whole anode thus rises by only slightly more than 50°. Even when several exposures are made in succession, no excessively high temperature results, especially since the dissipation of the heat developed is very much promoted by the oil in which the whole tube is immersed.

Permissible loading in fluoroscopy

In fluoroscopy, where the X-ray tube is more or less continuously loaded, the permissible power of the tube is almost entirely determined by the heat dissipation which in the previous section could be practically neglected. Therefore the fluoroscopy power depends not only on the tube itself, but also very much upon the construction of the container in which the tube is placed and on the method of cooling.

With forced oil cooling by means of a pump the small tube described could tolerate a continuous load of 1 to 2 kW. This is much more than ever occurs in medical fluoroscopy. Usually not more than about 180 W is necessary (namely in fluorscopy of the stomach which requires 85 kV, 3 mA), while, moreover, this load in only intermittent in most cases, with a ratio of approximately 1 : 2 between true loading time and intervals. The average (continuous) loading is therefore usually not higher than about 60 W. This makes it possible to employ forced oil-cooling and use the cooling by the natural flow of the oil inside the jacket, which simplifies the construction considerably. We shall now consider the problem of cooling somewat more closely.

The cooling must satisfy two main requirements: in the first place the anode lead of the X-ray tube may not become hot enough to constitute a danger to the glass, in the second place the temperature of the covering of the tube must remain so low that the user does not burn himself when he touches the tube. For the first requirement a maximum temperature of for instance 200 °C will be permissible, for the second the temperature must not be higher than about 60 °C, thus about 40° above room temperature. If the cooling is accomplished by means of oil, there is the additional requirement that the temperature of the oil must not rise far above 100 °C. It is known that at higher temperatures oil which is in contact with metals (especially copper and lead) rapidly depreciates, by acidification, formation of sediment, etc. with the result of a lower breakdown voltage and a poorer heat conduction [5]).

Let us first consider the second requirement. The heat dissipation of metal to air amounts to about 0.001 watt per cm² and per degree of temperature difference. In order to give off 60 W to the air with not more than 40° temperature dif-

Fig. 5. Jacket for the X-ray tube shown in figs. 2 and 3. The two large high-voltage leads may be seen, the window for the X-rays, to the left two rubber tubes for connecting the cooling spirals inside the jacket with the water main if desired, to the right the low-voltage cord for the automatic switching off of the high-voltage generator upon overloading.

5) In order to preserve the quality of the oil it is recommended in the literature that copper parts should be tin-plated. We have had very good results with chromium plating. Moreover, it is important to degas the oil well before use and not to allow it to come into contact with air or moisture during use. The latter is achieved satisfactorily in our case by means of the entirely closed jacket. Nevertheless it is advisable in such constructions not to have the electric field strength higher than about 50 kV/cm at any spot — a condition which in our case is also satisfied.

ference, therefore, the earthed jacket of the tube must have a surface of 1 500 cm². This can be achieved with a relatively small and therefore quite light jacket: the jacket shown in *fig. 5* made of brass sheet, together with tube and oil but without the cable, weighs only slightly more than 7 kilograms.

When in this way a temperature of not more than 60 °C is guaranteed for the jacket, provision may be made by suitable construction that the temperature of the flowing oil will never become appreciably higher, except in a thin layer close to the radiator. This radiator whose function is the transfer of heat from the anode block to the oil is in good contact with the anode and has a surface of about 200 cm². The heat dissipation of metal to oil for cases of natural flow in a relatively small space amounts to about 0.03 W per cm² and per degree temperature difference, according to measurements carried out in this laboratory. This value is valid for an oil temperature of about 60 °C; at a higher temperature the heat dissipation increases because of the fall in viscosity which permits more rapid flow. With this heat dissipation it may be calculated that for the dissipation of 60 W, the radiator must be about 10° warmer than the oil. We must, however, keep in mind that the power of 60 W represents an average load resulting from intermittent loading 1 : 2 with 180 W. We may here assume of course that the temperature of the flowing oil with its great heat capacity is determined only by the average load, and thus amounts to 60 °C; in and close to the radiator, however, where the temperature equilibrium is established much more quickly, the temperature during the working period must be calculated for a heat dissipation 180/60 = 3 times as large. In the working period, therefore, the temperature difference between radiator and oil is 30°, so that we find a temperature of 60 + 30 = 90 °C for the radiator.

This is the temperature averaged over the whole surface of the radiator. In this radiator which is 7.5 cm long, there is also a temperature drop. In our case for a dissipation of 180 W (working period) this drop amounts to about 20° so that the end of the radiator will be about 10° colder and the beginning about 10° warmer than the average. In the neighbourhood of the seal, therefore, a temperature of about 100 °C occurs which is indeed allowable for long times.

For those cases in which the dissipation of 180 W intermittent 1 : 2 is not sufficient, which may sometimes occur in very busy clinics, the jacket with the surface area of 1 500 cm² would not be able to dissipate the heat developed without exceeding the permissible temperature. In order to adapt the tube for 150 to 180 W continuous charge, a jacket with a surface area of about 4 000 cm² would have to be constructed. It is, however, also possible to continue to work with the convenient small jacket of fig. 5 if the oil in it is cooled. This can be done in the familiar way be introducing a fixed cooling spiral into the jacket, through which water from the mains is allowed to circulate when the higher load on the tube is employed. This is the solution we have chosen. The connection to the water main is by means of two thin rubber tubes along one of the high-voltage cables already present. For the rest the heat capacity of the whole is such that the tube can be used for about 15 minutes with 180 W without water cooling, beginning with the cold state. Only after the 15 minutes mentioned is the permissible limit of the temperature reached, and this may be seen on an indicator connected with an expansion box fastened to the jacket (the oil expands 0.07 percent per degree; see also the last article referred to in footnote [2]). If the load is too heavy, *i.e.* if the expansion box expands still farther, it automatically switches off the primary voltage of the high-voltage transformer.

Upon the use of water cooling the permissible continuous load is even considerably higher than 180 watts. By using a sufficiently high speed of flow of the cooling water the oil can be kept practically at room temperature (about 30 °C) even with powers of 300 W. For the temperature difference between radiator (middle) and oil we find about 50° for 300 W, for the temperature drop in the radiator about 34°, so that the anode seal need not become warmer than 30 + 50 + 17 = 97 °C. The temperature drop in the anode from the focus to the beginning of the radiator amounts to about 55°, the focus thus has a temperature of only about 150 °C, while temperatures up to 400 °C in continuous usage may be considered quite permissible. We may therefore develop a power of 300 W or more continuously in the tube with no disadvantage, and this is often desirable in the macroscopic examination of materials for which this tube may also be used.

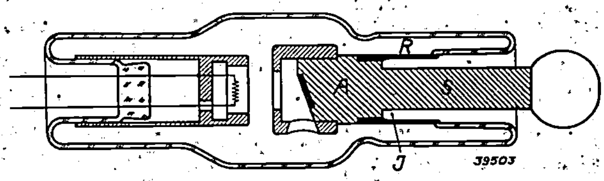


Fig. 6. Hard-glass tube for diagnosis of the type customary until lately, in which the anode seal *R* is welded into a depression in the tube.

These heavy continuous loads have been made possible here by the way in which the anode lead is fused into the tube. In the tube constructions common until now this weld was made in a concavity of the tube according to the sketch below (*fig. 6*), which is the simplest method for manufacture. Due, however, to the fact that the cooling body can only be relatively thin here and the flow of oil inside the depression *I* is very much hindered, a larger temperature difference between cooling body and oil and a greater temperature drop in the cooling body *S* occurs for a given heat dissipation, so that the seal and the oil become much warmer at that spot than is the case with the tube according to fig. 2 with the flat anode seal. In the case of such small tubes with this type of seal in a depression it is impossible to go higher than 150 watts continuous even with strong water cooling of the jacket.

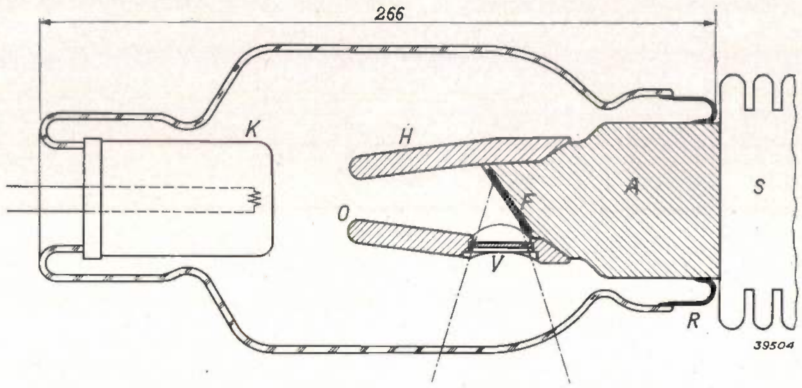


Fig. 7. X-ray tube for medical therapy. Apart from the larger dimensions the construction corresponds in general to that of the tube of fig. 2. The anode cap *H* is deeper, the opening *V* is closed with a beryllium window, the focus *F* is made very large 12×20 mm^2) for the sake of a low specific loading. Upon cooling the oil in which it is immersed by cooling spirals with running water the tube can tolerate a continuous load of about 3 kW (220 kV$_{max}$, 20 mA).

Tubes for therapy and material examination

In medical therapy as well as in the macroscopic investigation of the material of heavy pieces of work, considerably higher voltages, namely up to 200 and 300 kV$_{max}$, respectively, and with much higher continuous loads, several kW for example, are generally used. The type of tube which has been developed for these purposes is reproduced in *figs. 7* and *8*. In principle it is similar to the tube for diagnosis fig. 2, but larger in size. Because of the greater danger from secondary electrons at these voltages the anode cap *H* is made so deep that only about one per cent of the secondary electrons are able to fly back through the opening *O*. Furthermore, the opening *V* is closed to secondary electrons by a beryllium window 1 mm thick. The larger diameter of the tube (and thus also of the jacket) makes it desirable to have the undesired X-rays absorbed as far as possible by the anode cap in this case. The latter is therefore made of fairly thick tungsten copper so that the absorption in all directions corresponds at least to about 5 mm of lead The rays which pass through the opening *O* are absorbed by a tungsten plate on the cathode.

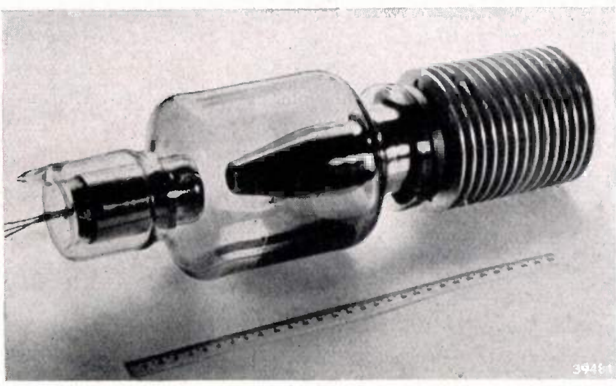


Fig. 8. Photograph of the tube for therapy whose cross section is shown in fig. 7.

In therapy, where it is not a question of obtaining an image, the focus of the X-ray tube need not be made especially small. For a loading by 220 kV pulsating voltage and about 20 mA (about 3 kW), which is now considered in therapy to be a high power, we have chosen a focus of 12×20 mm^2, with which the specific focal loading amounts to only about 12 W/mm^2.

For the continuous dissipation of the high power mentioned use has usually been made of forced oil or water cooling of the anode. Due to the small surface of the anode lead the liquid must here be sent through relatively narrow openings, which when oil is used sometimes leads to too high oil temperatures with the result of deposition of carbon and clogging. When water is used this no longer holds, but a circulation pump must be used which is insulated for high voltage, and the high-voltage cable for the anode must be constructed with an axial channel for the water circulation, which again introduces other practical difficulties. When it is also kept in mind that the pumps cannot usually be made noiseless, it is clear that it is a great advantage to be able to do without forced cooling.

We have indeed succeeded in this aim by the application of the principles already described in connection with the tube for diagnosis, but with larger dimensions than there used. A radiator with a surface area of about 1 500 cm² is connected to the anode lead and around this radiator, inside the oil-filled jacket, water-cooling spirals are placed at 4 to 6 cm distance from the radiator, see *fig. 8*. These spirals are earthed and may therefore be conncoted with the water main. With this arrangement, by means of the natural flow of the oil, 3 kW can be dissipated continuously without the anode or the oil becoming too hot.

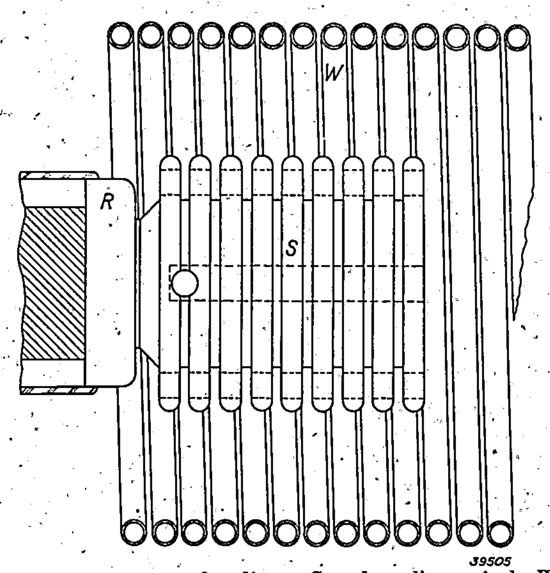


Fig. 9. Arrangement of radiator *S* and cooling spirals *W* in the high-power tubes.

The tube drawn in fig. 7 can be operated on 220 kV A.C. voltage and about 300 kV pulsating D.C. voltage. A tube of the same type can also be constructed with earthed anode and the anode cooled directly with water from the main, so that appreciably greater powers can be dissipated than mentioned above.

In tubes for material testing the focus is naturally made smaller than mentioned above for therapy tubes. In general the apparent dimensions of the focus may not be much larger here than 4 mm square. Then with the radiator cooling described powers of more than 1 kW can be dissipated continuously without the temperature of the anode exceeding the permissible value.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

1537: A. Claassen and J. Visser: The volumetric determination of titanium (Rec. Trav. chim. Pays Bas **60**, 213-223, Mar. 1941). (Original in German).

If provision is made for the entire absence of oxygen in the solutions, especially in the titration liquid, in the titration of trivalent titanium with solutions of ferric salts, results can be obtained which are accurate to within 0.1 per cent, if one simply uses the theoretical iron content of the titration liquid. Too low values are, however, found if the titration is carried out at temperatures above 40 °C. It is advisable to use the cadmium reducer proposed by Treadwell for reducing the titanium. In order to determine iron volumetrically in addition to titanium it is best to apply the reduction in the silver reducer.

1538: K. F. Niessen: On the acoustic analogue of Sommerfeld's surface wave (Physica **8**, 337-343, Mar. 1941). (Original in German).

If the sound wave of a loud speaker is propagated in the neighbourhood of a flat surface of a porous material, the wave front of the sound always dips forward, no matter how close the loud speaker is placed to the surface. This must not be explained in the experiments of Janovsky and Spandöck discussed in this article on the basis of the acoustic analogue to the space radiation according to Sommerfeld, but with the help of the analogy with its surface wave. Use should here be made of the analogy indicated by Schuster of the dielectric constant and the conductivity.

1539: J. L. Snoek: The determining factors of permeability (Physica **8**, 344-346, Mar. 1941).

The concept defended in 1936 by the author (*cf.* **1078**), that for homogeneous annealed alloys the permeability depends only upon the internal stresses and the crystal anisotropy, has since been confirmed by investigations of Grabbe on slowly cooled nickel-iron alloys, and of Williams and Bozorth on single crystals of iron and silicon-iron. In contrast to this, however, the experiments which Snoek and Rathenau recently carried out on cold worked nickel-iron alloys do not show the least agreement with this concept. It is highly probable that some unknown factor is here involved.

1540: J. F. Schouten: De experimenten van Seebeck met de sirene en de acoustische wet van Ohm (Ned. T. Natuurk. **8**, 154-165, Apr. 1941) (Seebeck's experiments with the siren and Ohm's acoustic law).

Various investigations on the perception of sounds which were carried out by Seebeck with his siren in 1841 and which at that time led to a lively discussion, since according to him they could not be explained by the acoustic law then proposed by Ohm, were repeated by the writer and confirmed. It is found that on the basis of modern insight into subjective sound analysis (*cf.* Philips techn. Rev. **5**, 286, 1940) these phenomena can easily be explained and that the explanation lies in the direction in which Seebeck sought it 100 years ago.

1541: K. F. Niessen: On the testing of Pauling's hypothesis about the binding of the atoms in metals (Physica **8**, 377-386, Apr. 1941). (Original in German).

An attempt is made to find a confirmation of Pauling's hypothesis about the binding of the atoms in the metals scandium to copper from the measured Curie temperature of nickel. This attempt is unsuccessful no matter how the number of 3d electrons prescribed by Pauling are distributed among the levels of the ferromagnetic electron band.

1542: P. J. Bouma: Physiological optical foundations of the problems of air-raid protection blackout (Physica **8**, 398-412, Apr. 1941). (Original in German).

For the main points in the contents of this article the reader is referred to Philips techn. Rev. **6**, 161, June 1941. Furthermore on the basis of new experimental data nomograms have been constructed from which the relation can be read off between the brightness at which a light spot can just be observed and the angle of vision at which it is seen.

1543: P. J. Bouma: The relation between the concepts Brightness and "Dunkelleuchtdichte", etc. (Physica **8**, 413-423, Apr. 1941). (Original in German).

For the contents of this article the reader is referred to Philips techn. Rev. **6**, 161, 1941.

1544: M. J. O. Strutt and A. van der Ziel: What quantities characterize the suitability of an electronic tube for the amplification of the weakest signals? (Physica **8**, 424-425, Apr. 1941). (Original in German).

In this article emphasis is laid on the fact that the ratio of the input noise resistance to the input resistance between control grid and cathode is not a suitable quantity for the characterization of the usefulness of an amplifier valve for the amplification of small signals. More suitable for this purpose is the ratio between that part of the noise resistance which is a result of fluctuations in the cathode current and that part of the input resistance which is a result of the transit times of the electrons in the valve (*cf.* Philips techn. Rev. **6**, 178, June 1941).

VOL. 6. No. 11, p.p. 317-348 NOVEMBER 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

TECHNICAL PROBLEMS IN THE CONSTRUCTION OF RADIO VALVES

by TH. P. TROMP. 621.385

In this article several episodes in the history of the development of radio valves are discussed, particularly those concerned with the form of the valve. It is shown how the technical problem of whether glass or metal is the more suitable material for the bulb of the valve led to a complete revision of valve construction. The result is that radio valves can now be made either of glass or metal which satisfy all the requirements made by ordinary sets for the broadcasting region. The Philips key valves in glass or metal are finally discussed in detail. The conclusion is drawn that the difference in properties for the broadcasting region is so slight that neither of the two models is essentially preferable to the other. In the region of metre waves and shorter, however, the all-glass valves exhibit increasing advantages with decreasing wave length.

Introduction

The development which is observed in the form of every technical product, be it an electrometer, a lamp or an amplifier valve, is in general directed by two factors: first an ever-increasing insight into the physical laws which determine the function of a product, and second the problems and possibilities of manufacture. It may in general be said that for products which are manufactured only in small quantities the first factor is the more important, while in mass production the second factor also comes into prominence.

The question of which of the two above-mentioned factors determines the technical development also depends upon the nature of the product. If one considers a radio valve it may be said that the technical development of its electrode system is directed by a steadily increasing insight into the physical phenomena which determine its function. The external form of the radio valve, however, is quite a different question. As far as the function of the valve is concerned, the envelope, aside from providing contact in the socket, serves only to maintain a vacuum around the electrodes. From this point of view the external form of a radio valve would logically be a container which encloses the electrode system as closely as possible and through which mutually insulated connections are led. These connections would preferably be chosen in the form of short, straight wires, strong enough to be used as contact pins and far enough apart to cause no undesired couplings between the various electrodes.

In this article we shall give several details from the history of the development of radio valves which have indeed finally led to constructions of the simple character described [1]). The fact that this has occurred only after many years may not be ascribed to a lack of insight, but only to the difficulties which occur in the manufacture of a vacuum-tight container with sealed-in connections.

Historical survey

It would lead us too far if, in this article, we were to go into all the problems which are here encountered. It may be stated simply that the problem of the manufacture of a vacuum bulb with leads had previously reached an almost perfect solution in lamp manufacture, in the form of the so-called "pinch", so that it was obvious that this principle of construction could also be used for radio valves. The valves with pinch construction consist entirely of glass and are provided with a base with contact pins. The connecting wires to the electrodes must be soldered into these pins which are moulded or pressed in.

Although this construction by no means satisfies the above-mentioned requirement of smallest possible dimensions, and although the use of a base with pins to which leads must be soldered meant an undesired complication, this construction was

[1]) See Philips techn. Rev. 4, 162, 1939.

able to maintain its supremacy until the use of new materials for the envelope of the valve introduced a complete change in the form.

In 1935 in America metal radio valves appeared on the market (G.E.Co. and R.C.A.) [2]). Relatively soon after this, metal valves also appeared to a limited extent in some European countries. Philips manufactured a large number of such valves for European and overseas markets.

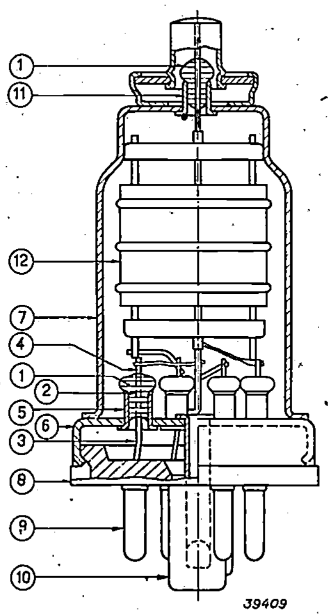


Fig. 1. Radio valve with metal envelope which appeared on the market in America in 1935.
1. Glass bead,
2. Molybdenum lead,
3. Nickel or copper connector,
4. Nickel support wire,
5. "Fernico" eyelet,
6. Metal header,
7. Metal bulb,
8. Bakelite base,
9. Hollow contact pin,
10. Bakelite spigot,
11. Eyelet of top lead,
12. Electrode system.

The first metal construction consisted of a fairly complicated arrangement, of which *fig. 1* illustrates a cross section. Briefly it is as follows. The "header" of the valve consists of a horizontal metal plate with a vertical cylindrical edge. In this header a number of holes are punched in which "Fernico" eyelets are welded. "Fernico" is an iron-nickel-cobalt alloy which can easily be welded to hard glass with a comparatively low coefficient of expansion [3]) — about 50 $\times$ 10^{-7}/°C. After welding, the eyelets are soldered with copper in order to obtain an absolutely reliable vacuum-tight joint. There is also a hole in the centre of the header in which a metal exhaust tube with a small flange is welded for the evacuation of the valve later on.

Lead-in wires are sealed into the "Fernico" eyelets (coefficient of expansion 48 to 50 $\times$ 10^{-7}/°C) with the help of the above-mentioned hard glass with low coefficient of expansion. These leads consist of three parts: an internal nickel support for welding the electrodes of the system, a short section of molybdenum wire for the vacuum-tight lead-in and a nickel or copper wire for the connection to the pins of the base, which in this case is constructed as a bakelite "wafer type" base plate with pressed-in pins. In this case a "holder" is therefore not superfluous, nor is the soldering of the leads in the pins. From the electrical point of view also this is a disadvantage, for instance with the high frequency losses and the frequency-drift during the heating-up period.

After the leads have been sealed in, a series of manipulations follows before the electrode system can be welded on the header, and only after that has been done can the iron envelope be welded to the base plate by means of a very heavy projection welder [4]) which produces a vacuum-tight weld. The valve is then ready to be exhausted.

It may be noted that the protection of the exhaust tube which is squeezed, welded and cut off, after the exhausting and de-gassing, is provided by a central spigot which at the same time locates the valve in the socket correctly.

A variant of this model, of which it was hoped that it would be less expensive than the construction described above, has the metal header replaced by a glass plate with sealed-in lead wires (nickel-iron core with copper covering). The glass plate is sealed into a chrome-iron ring which in turn is welded into an iron ring. The iron ring has the same function as the edge of the header of the model previously described, namely for the bulb to header weld. In this model the exhaust tube is not of metal but of glass; see *fig. 2*.

Originally in this model of valve the control-grid lead-out was at the top of the metal bulb as shown in the figure.

These valves were appreciably smaller than the glass valves previously manufactured with pinch construction, and due to the metal envelope a good

[2]) We do not consider here constructions in which the bulb serves not only as envelope but at the same time forms a part of the electrode system. Such constructions have long been used in high power transmitter valves, where the anode is constructed as a part of the envelope in order that it may be cooled more easily. In the case of receiver valves also such a construction, the so-called "Catkin" valve, was designed in England. This valve was, however, not manufactured on a large scale and after some time it disappeared from the market.

[3]) The coefficients of expansion of glass mentioned in this article refer to average values which are measured in the temperature range between 20 and 320 °C.

[4]) A projection welding apparatus is a resistance welding apparatus with which the parts to be welded can be joined on an accurately defined surface by a welding impulse of very short duration and considerable power.

shielding of the electrode system against external fields was automatically obtained.

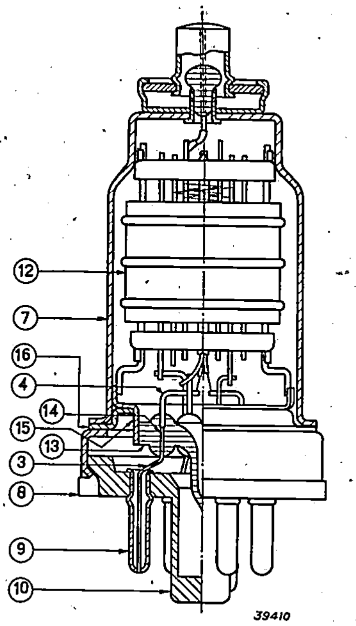


Fig. 2. Radio valve with metal envelope in which the metal header is replaced by a glass "button stem". For the meaning of the numbers see fig. 1, and further:
13. Glass button stem,
14. Lead-in wire,
15. Chrome-iron ring,
16. Iron ring.

In the meantime a new form of metal valve appeared in the European market (developed by "Telefunken"), in which the experience gained with the American construction was taken into account and several new ideas incorporated (see *fig. 3*). The principle of the metal header with "Fernico" eyelets and the insulated molybdenum leads through hard glass were here again used; a "base" in the form of a flat plate with moulded or pressed-in contact pins must also be used under this metal valve. The difference between this valve and the American model according to fig. 1 is that the electrode system is constructed h o r i z o n t a l l y, which permits a very solid construction, but with the accompanying (not inconsiderable) increase of the diameter. An advantage is that all the electrodes can have their leads on one side, namely to the bottom and without thereby increasing the anode-grid capacity [5]) (single-ended construction). In a series of valves this capacity is very important, since it may be the cause of an undesired coupling of anode and grid circuit. In modern set construction it is required that this capacity shall not exceed the extremely small value of 0.002 pF, a requirement which can immediately be fulfilled with the new form of metal valve.

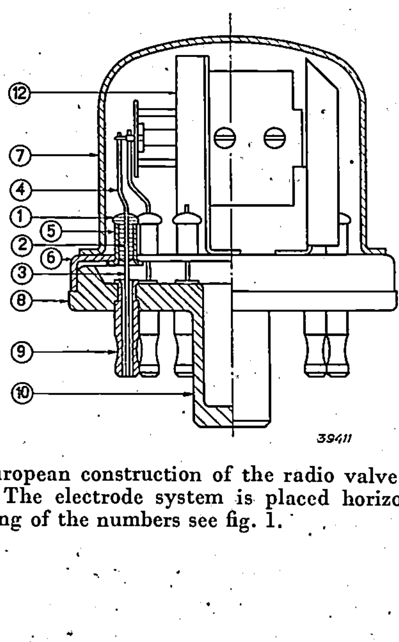


Fig. 3. European construction of the radio valve with metal envelope. The electrode system is placed horizontally. For the meaning of the numbers see fig. 1.

Technically this construction is satisfactory in all respects, it has, however, various disadvantages, for example the fairly large diameter and the fact that power-output valves and rectifier valves of the power as used in the broadcast-field cannot be made according to this design because of the limited length of the cathode (due to the horizontal construction). For these types of valves a vertical construction would again be necessary. It is for this reason that this metal range consists chiefly of radio-frequency, frequency-changing, intermediate-frequency, detector and low-frequency amplifier valves, while output valves and rectifiers belonging to this series are manufactured in the old familiar pinch technique, but with the same base as the metal valves.

The all-glass key valves

In the meantime work has been done at Philips on an improvement in the construction of the glass valve. The requirements upon which this improvement was based are briefly the following:
Small dimensions,
Good shielding,
Reproducibility in manufacture,
Good contact and firm fastening of the valve in the valve holder,
Low consumption of material and preferably use of materials which can be easily obtained.

In order to simplify the manufacture it is furthermore desirable that the valve should have no base, but the leads should be in one piece and serve also as contact pins.

[5]) Later this design with all connections on the same side was also applied in America to the constructions represented in figs. 1 and 2. This was done by bringing the grid lead from the top to the bottom and introducing extra shielding in order to prevent the anode-grid capacity from becoming too large. This capacity is, however, much larger than that of the European metal valve construction described.

The result was the modern all-glass valve construction described in the article referred to [1]). It is chiefly distinguished by the use of pressed glass and thick chrome-iron contact pins, which make it possible to omit the "holder" which is in many respects so undesirable. In connection with the form and the function of the spigot these valves are called key valves (see *fig. 4*).

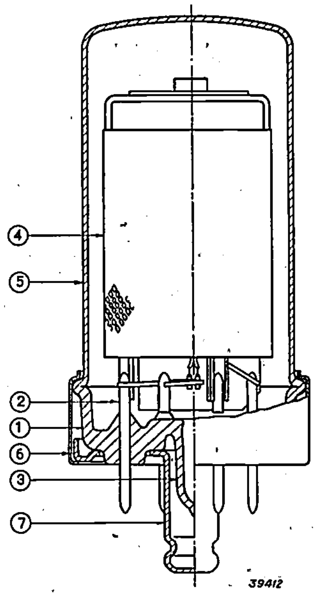


Fig. 4. Cross section of an all-glass Philips key valve. The sealed-in pins make direct contact in the socket; the bakelite base present in previous valves is eliminated.
1. Pressed glass base,
2. Chrome-iron pin,
3. Glass exhaust tube,
4. Electrode system,
5. Glass bulb,
6. Metal ring,
7. Metal spigot.

The model in which originally a series of special valves (particularly for television and other special purposes) were manufactured, still had a rather large diameter (envelope about 30 mm; largest diameter about 34 mm), but the use of these valves already meant a considerable saving of space on the surface of the chassis compared with the use of the metal valves with a diameter of 43 mm.

From this model with nine pins the modern glass key valve was developed with eight pins, an envelope diameter of about $27^1/_2$ mm and a largest diameter of about $31^1/_2$ mm. The central spigot in these valves is of metal and serves not only as centering and locating device in inserting the valve into its socket, but at the same time as shielding between anode and control-grid pin. This spigot is so arranged that the valve clicks into the socket, which is a great advantage for the locating device if the set is to be transported complete with valves; a requirement which at present is universal.

This construction also makes it possible to obtain a grid-anode capacity which is smaller than 0.002 pF in spite of the vertical assembly, the small diameter and the single-ended construction. The object of making smaller and smaller valves in connection with the demands of the set-makers is fully met by this construction. The development of the Philips midget receiver, type 203 U, which although equipped with four valves, has a volume of only 5 dm³, was only possible because of the existence of such valves.

Properties of glass and metal valves

The advantages of the all-glass valve compared with the valves with a pinch have already been discussed in detail in the article cited [1]), and it is these properties which make them particularly suitable for use in the region of ultra short waves. The shorter the wave length, the greater the advantages. In the televison region (wave lengths of several metres) the properties are already important, at still shorter wave lengths (decimetre waves) they may even become of such value that a pinch construction need no longer even be considered.

We shall now compare several properties of the glass key valves with those of metal valves.

During the heating up period of the valves in a set, the capacity variations which occur due to the change in the dielectric cause a frequency drift of the set which has previously been pointed out. At broadcasting wave lengths this frequency variation may already be somewhat disturbing; but with the decrease in the wave length it becomes steadily more important. While with the glass key valves a decrease of the frequency variation by a factor 2 (at 15 m wave length) compared with the pinch construction was obtained, with the metal valves this advantage is again lost.

A second disadvantage of the metal valves is that in the construction described the leads sealed into the eyelets involve a rather large concentrated capacity, whereby at wave lengths shorter than 1 m an appreciable depreciation in the results, which can be obtained with the valves is noticeable, and below about 50 cm even makes their use impossible.

A third point of difference relates to the input and output damping. In the case of the glass valves this is lower than in the metal construction, since the dielectric losses of the capacities between the leads and the metal envelope furnish an extra contribution to this damping. Because of this a greater amplification can be obtained with glass valves at wave lengths in the metre region.

It is clear that all these points of difference relate to the short-wave properties. In the case of the ordinary broadcasting receivers, however, they are of no great importance.

As a matter of fact the metal valve has the advantage already mentioned of providing a better shielding against external fields. In the case of the glass valve this shielding must be obtained by building in a so-called cage of metal gauze (at least for those types where shielding is necessary), which encloses the whole system — a device which is satisfactory in all respects.

Another important point, which is of importance not only in the glass but also in the metal construction, is that of the high-frequency losses of the leads in the ultra short wave region. The D.C. resistance of the solid pin is hereby of secondary importance, because the high-frequency resistance depends mainly upon the condition of the surface layer. With given dimensions of the lead the following values were measured at a wave length of 90 cm:

silver wire	0.12 Ω
copper wire	0.12 ,,
aluminized iron	0.20 ,,
molybdenum	0.23 ,,
tungsten	0.26 ,,
A grade poorer are:	
nickel	1.20 ,,
iron	3.60 ,,
"Fernico"	5.20 ,,
chrome-iron	6.00 ,,

From this it will be clear that if pins of the last mentioned materials are copper-plated or silver-plated, considerable improvement can be obtained. Thus, for example, with a given lead-in, thin silver-plating of the chrome-iron wire gave an improvement of the high-frequency resistance from 1.1 Ω to 0.13 Ω for $\lambda = 1.0$ m, while in another case an improvement from 1 Ω to 0.17 Ω was obtained by copper-plating the wire.

A metal key valve

From the above considerations it follows that in principle it must be possible to design a metal valve for the broadcasting field which satisfies the requirements specified in the introduction. Such a valve may be considered as equivalent in practical use to the glass key valve construction described. It is possible therefore to be more or less independent of the supply of a given raw material, factory equipment, etc. The solution of this problem is shown in the valve construction given in *fig. 5.* As to dimensions this valve is identical with the all-glass key valve described: the contact pins are also in the same relative positions to each other, there is also a spigot, its over-all dimensions are the same, and the same electrode systems can be used in it; briefly it is a design with which it is possible to make replicas of existing all-glass types. The components of this type are shown in *fig. 6.* The capacities between the electrodes (for instance input and output capacities) will of course differ slightly from those of a similar type in all-glass construction, but these differences are so small that they can be compensated very easily by trimming the set. In many cases it will even be possible to replace the valves in existing sets without any change.

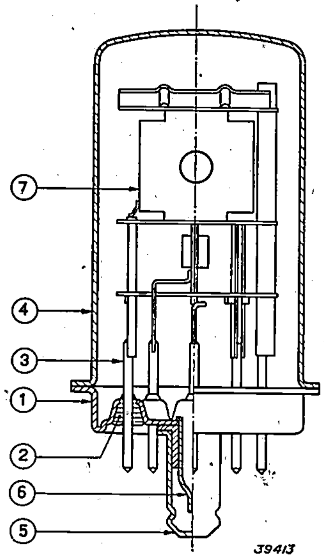


Fig. 5. Cross section of a metal key valve. The external dimensions, the arrangement of pins and the construction of the electrode system are exactly the same as in the glass key valves.
1. Metal header,
2. Glass bead,
3. Iron pin,
4. Metal bulb.
5. Metal spigot,
6. Metal exhaust tube,
7. Electrode system.

The new metal valve consists in principle also of a header (see figs. 5 and 6) with small holes for the vacuum-tight sealed-in leads, while there is an opening in the centre for welding the exhaust tube into the header. The advantage of this construction compared with that according to fig. 3 is, that due to the strong vertical construction in which, just as in the all-glass valves, several U-shaped supports provide the necessary mechanical strength, the length of the cathode may be chosen in accordance with the type of valve to be made. Power-output valves, rectifier valves, special combination valves, etc. can therefore also be made in this construction.

The iron header is subjected to several special treatments before the eight leads are sealed-in at a temperature of about 950 °C. A glass bead is previously fused around each pin, the glass of which is adapted to the lead wire and the iron header. The pins may then be of ordinary iron or of an iron alloy. In both cases they are subjected to a special surface treatment before sealing-in.

Upon comparison with the previously described metal construction it is obvious that the quality of this seal must be very high, not only with regard to vacuum tightness, but also to that of mechanical strength, accuracy and precision. The electrode system of the valve must be welded to the pins on expansion, *viz.* about 103 × 10^{-7}, and it is fused into a soft glass with a slightly lower coefficient of expansion. This glass has a much lower softening temperature, namely about 475 °C.

Since this chrome-iron to glass seal satisfies

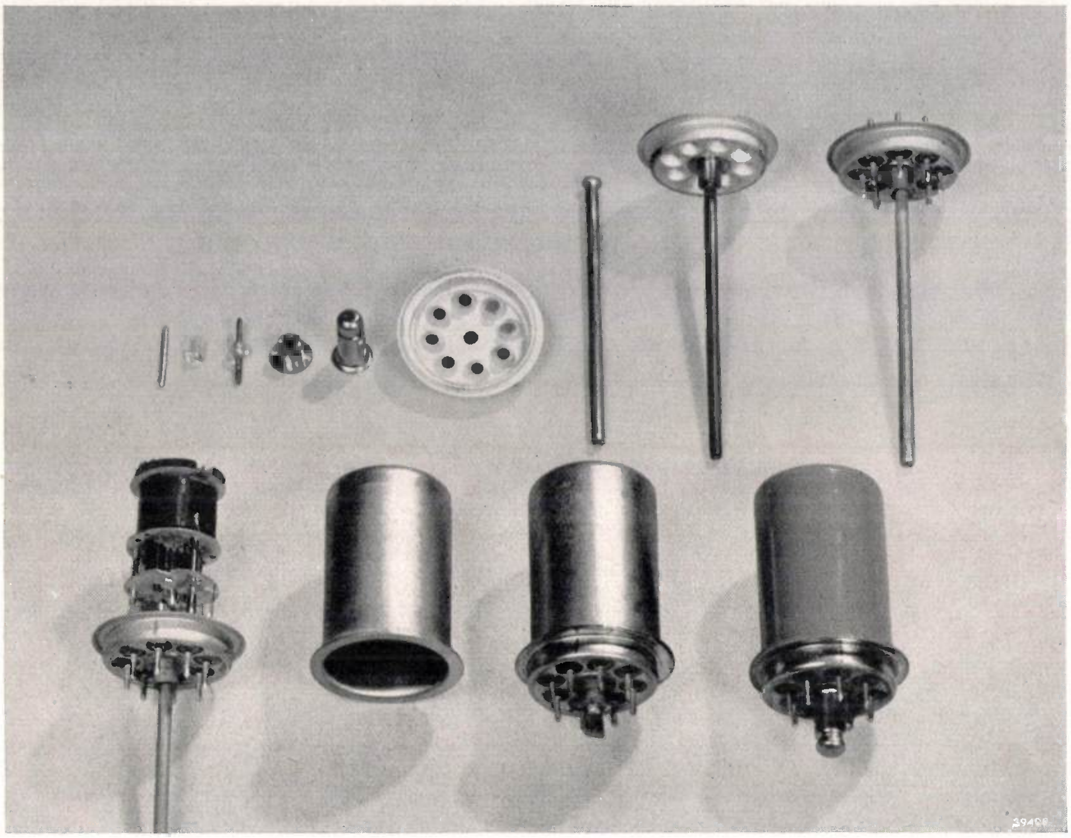


Fig. 6. The parts of the metal key valve in different stages of manufacture. Upper row, left to right: one of the 8 fused-in leads; glass tube to make bead; bead fused on pin; "cap" upon which the spigot is later welded; spigot; metal base plate with holes; metal exhaust tube; combination of base plate, exhaust tube and cap complete the base plate with fused-in pins. Lower row: electrode system assembled and mounted on the pins; metal envelope; valve after being exhausted and finally the completed valve.

the inside of the valve, and on the outside of the valve (when inserting it in its socket) fairly strong forces are exerted on the pins. The technical problems which are encountered in making those leads are not simple and require a thorough knowledge of glass-metal seals and wide experience. A brief comparison of the different techniques used in the constructions previously described gives the following picture.

In the case of the sealing in of a thin molybdenum wire (coefficient of expansion 51 × 10^{-7}) to give a vacuum-tight seal and to be insulated in a "Fernico" eyelet (coefficient of expansion 48 × 10^{-7}) a glass is used with a coefficient of expansion as close as possible to the above. A hard glass has been used with a softening temperature of about 540 °C [6]).

In sealing the solid chrome-iron pins in the glass key valve, the pin has a much higher coefficient of high requirements as to quality and reproducibility, it would seem obvious that the header of the metal key valve as well as the pins should be made of chrome-iron. This would indeed be excellent were it not for the fact that the cost of such a header would be too high for normal use. Therefore an attempt was made to use cheaper and more easily obtainable metal. It was for this reason that iron was chosen for the new construction, with its appreciably higher coefficient of expansion: 125 × 10^{-7}. This makes it necessary that the glass should again have a higher coefficient of expansion.

Although in general soft glass is easier to work

[6]) The softening point is defined as the temperature at which a glass rod 30 cm long and 4 mm in diameter, supported at each end, bends 2 mm under a weight of 195.5 g. The melting point is much higher.

with than hard glass, soft glasses with such high coefficients of expansion introduce difficulties in manufacture among which is the chemical decomposition of these lead glasses during the life of the valve, not only by atmospheric influences but also by the formation of "lead trees" due to electrolysis of the glass. Moreover, in the short period during the sealing-process (at a temperature of about 950 °C) reduction of the lead oxide may occur resulting in surface conduction.

A satisfactory solution of these problems has been found and a manufacturing process worked out for the surface treatment of the metals and the sealing-process which meets these objections.

It is also important that the construction should be such that as far as possible the glass will be under pressure stresses only and under all conditions of use, since only under these conditions can an adequate guarantee be obtained of a mechanically strong and at the same time vacuum-tight seal (no cracks, no leaks).

When we study the stress diagram of such a complete header it will usually be evident that only pressure stresses do occur in the seal of glass to base plate, and at the pin, axial pressure and tangential pressure is combined with radial tensile stress. It may, however, also occur that the stress diagram of the axial and radial stresses along the pin is reversed (axial tension and radial pressure stress). The whole picture becomes more complicated when it is kept in mind that variations in the coefficient of expansion of the glass are inevitable and that the wire and the header do not behave in the same way, due to the fact that the influence of the surface layer on base plate and pin is relatively different.

After the metal exhaust tube has been welded on the header and the pins have been sealed in, the header is nickel-plated. On the bottom of the header a small "cap" is welded which is used afterwards for the fastening of the spigot which also serves to protect the iron exhaust tube. The metal bulb is then welded on the header with a heavy projection welding apparatus, after the electrode system of the valve has been mounted and welded to the pins and the connections between the various electrodes and pins have been made.

The following figures will give some idea of the power needed for such welding. While for the previously described metal valve with a diameter of 43 mm a projection welding apparatus of nominally 400 kVA must be used (the power impulse during welding is actually of the order of magnitude of 1000 kVA; primary 500 V, 2000 A, secondary about 110,000 A), for this construction a smaller apparatus can be used, of nominally 175 kVA (peak about 450 kVA). These welding machines are regulated by a special control device which makes it possible to vary the duration of the weld impulse continously from about $^1/_2$ to about 25 cycles, and to set the welding current in 16 different positions by primary regulation. The welding pressure can also be regulated between 100 and 17,000 kg. The welding electrodes are developed for this purpose and are water-cooled.

The welding method described has the advantage over any sealing process that the attachment of the envelope to the base header takes place at room temperature (it is only necessary to dissipate the heat developed at the weld and this is relatively little due to the short duration of the weld impulse). Undesired oxidations in the inside of the valve are thus *a priori* avoided. In the case of glass sealing, on the other hand, in certain cases precautions must be taken such as the protecting of the interior of the valve with inactive gases like nitrogen, or mixed gas.

The valve thus constructed can be exhausted and de-gassed in the ordinary way. The exhaust tube is then squeezed and sealed by welding.

It is obvious that in the case of a metal valve direct heating of the internal parts by means of high frequency is impossible. The heating of the internal parts for de-gassing purposes must therefore take place indirectly *via* the metal envelope or by partial heating from the inside by heating the filament, with or without the combination of electron bombardment of the grids, etc. For external heating of the envelope (the header and leads must of course be cooled at the same time) high-frequency heating is, however, unnecessary, and the heating can more economically be done locally by means of gas.

The getter for binding the gas residues can be very easily flashed by heating with gas flames from the outside, or by the passage of current, and it is then deposited on the wall of the envelope with the advantage that the layer of getter is connected to zero potential (and thus does not form a "floating" capacity).

Summarizing, it may be said that it is possible in the broadcasting region to eliminate the problem of a choice between "metal" or "glass", so that the existence of this alternative need have no effect on the further development of set construction. Only considerations of a practical nature (equipment present in a given factory, the availability of certain materials or other special circumstances) need in the future determine the choice of the technique to be employed.

ASSEMBLING GRIDS FOR RADIO VALVES

In triodes the interior of the valves consists of a cathode with a grid and an anode. Two grids are used in tetrodes, pentodes require even three, and so on.
Reproducibility and interchangeability of valves, and their characteristics, necessitates that the interspacing of these parts and their dimensions be always the same within very narrow limits. The assembling operations, consisting mainly in welding and pinching, are therefore onerous and precise and require many checks and random tests.

CARRIER-WAVE TELEPHONY

by D. GOEDHART and J. de JONG.

621.395.44

In the case of telephone connections over long distances, where the cable and the repeaters connected to it represent an important part of the total expense, it is often advantageous to multiply the number of conversations which can take place simultaneously on one set of wires by the application of carrier-wave telephony. The principles of carrier-wave telephony and the general construction of an installation for this purpose are described in this article. From the functions of the different components the requirements which must be made of them are deduced, while the specific problems and the practical construction of the parts will be dealt with in future articles.

Introduction

The simplest way of bringing about a telephone connection between two subscribers A and B is to connect the microphone of A by means of two wires to the telephone of B, and the microphone of B by two other wires to the telephone of A. Four connection wires are hence needed (four-wire connection, *fig. 1*).

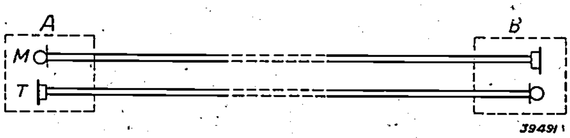


Fig. 1. Four-wire connection between two subscribers A and B.

Connections were very quickly worked out in which the number of connecting wires could be reduced by one half (two-wire connection, *fig. 2*). In each of the two telephone instruments there is then a so-called fork, an arrangement which sends the microphone currents over the line and the currents incoming from the line to the telephone.

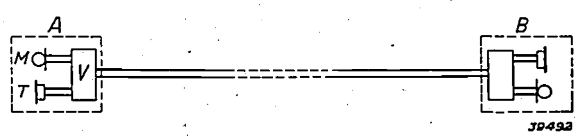


Fig. 2. Two-wire connection between two subscribers A and B. V forks.

Due to the damping caused by the connection wires the distance which can be linked in this way is limited. After the invention of amplifier valves, amplifiers (repeaters) could be placed in the connection line [1]), thereby making the distance covered many times greater (*fig. 3*). With this increase in the distance the connection line becomes proportionally more expensive, while the costs of the terminal apparatus (microphone, telephone, fork) remain the same, so that the greatest part of the cost of the connection is due to the connection line.

If many calls must be transmitted simultaneously between two centres of population, the same number of pairs of connection wires with the necessary repeaters are required. These connection wires are usually combined into cables with many cores. By the application of carrier-wave telephony, however, different conversations can be carried on at the same time over one pair of connection wires. This makes the cost of the cable and the repeaters much lower. The apparatus at the beginning and end of the connection also becomes more complicated and therefore more expensive. From this it follows that the application of carrier-wave telephony is only justifiable economically for the bridging of long distances where the cable costs are the dominating factor; for small distances, for instance for local calls, the cost of the final apparatus would be too high and carrier-wave telephony will therefore not be used.

In two previous articles in this periodical [1,2]) several problems of carrier-wave telephony have already been discussed. In this article we shall consider the general plan of an installation for carrier-wave telephony and the most important problems which occur in such an installation. The solutions

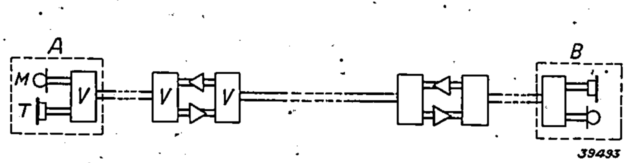


Fig. 3. If the distance between two subscribers A and B is great, repeaters are inserted in the line at regular intervals. Since repeaters give transmission in only one direction there are always two side by side, one for each direction, with a fork at each side in the case of a two-wire connection.

[1]) We do not here consider coil-loading by which means the distances bridged are also made greater. The subject is discussed in detail in: F. de Fremery and G. J. Levenbach, Carrier-wave telephony on coil-loaded cables, Philips techn. Rev. 4, 20, 1939.

[2]) G. J. Levenbach and H. van de Weg, Non-linear distortion in loaded cables, Philips techn. Rev. 4, 79, 1939.

which have been found for the problems encountered, and the construction of the necessary parts will be discussed in following articles.

The principle of carrier-wave telephony

The vibrations occurring in speech may be considered as made up of a large number of sinusoidal vibrations with different frequencies and different amplitudes (actually an infinite number of vibrations, represented by a Fourier integral). For a satisfactorily intelligible conversation it is not necessary to transmit all these frequencies, but only those which lie between about 300 c/s and about 3000 c/s. The telephone cable can, however, transmit a much wider frequency region. The frequency region above 3000 c/s can now also be used for the transmission of telephone conversations by the use of a method which exhibits much analogy with that of radio technology.

A carrier-wave with a high frequency f_c is modulated in a transmitter with the low-frequency oscillations of speech. Due to this modulation two new groups of frequencies occur: the so-called side bands at either side of the carrier wave (*fig. 4*). The

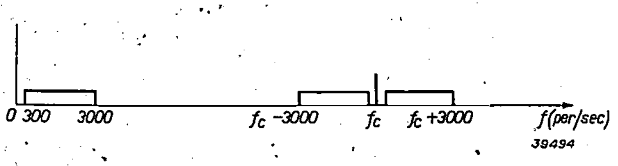


Fig. 4. Position in the frequency spectrum of the speech vibrations to be transmitted and of the carrier wave with two side bands.

upper side band (from $f_c + 300$ c/s to $f_c + 3000$ c/s) represents a shift of the frequency spectrum of 300 to 3000 c/s by an amount f_c, the lower side band (from $f_c - 3000$ c/s to $f_c - 3000$ c/s) is the mirror image of the upper side band with respect to the carrier-wave f_c. In this way a number of different conversations can be modulated, each on a carrier wave with a different frequency and all these modulated carrier waves can be fed to the common line.

If the frequency difference between the carrier waves is sufficiently great, then with a corresponding number of selective receivers connected at the other end of the line, each of the modulated carrier waves can be received without interference by the others. Just as in radio technology the ether transmits various programmes at the same time, in the same way in carrier-wave telephony a single pair of cores simultaneously transmits many conversations. It is clear that the efficiency of the cable is increased many times in this way.

If the same pair of cores is used for the transmission of conversations in both directions (two-wire connection, fig. 3) several complications result, since it is difficult to prevent mutual interference of the speech in opposite directions (see article [1]). Therefore the four-wire connection (fig. 1) is usually used in carrier-wave telephony. This, however, does not require more pairs of cores than the two-wire connection. It is found, that in the two-wire connection it is not easily possible to use the same carrier-wave frequencies for the two directions, because of the difficulties which the forks needed at each side of each repeater occasion at high frequencies. Each frequency region can therefore only be used once in a given pair of cores, either for one direction or for the other. The number of calls is then the same as if one half of the pairs of cores were used exclusively for one direction and the other half for the other direction (four-wire connection). Ordinarily two cables are laid side by side with equal number of pairs of cores, one cable for each direction.

It is not necessary to transmit both the side bands formed in the transmitter; one side band is enough. The receiver is capable of reproducing from the carrier wave and one side band the low-frequency vibrations which formed the original speech. With the help of a filter, therefore, at the transmitting end one of the side bands is cut off, which makes it possible to include in a given frequency region twice as many call channels, and the efficiency of the cable is thus doubled (*fig. 5*).

As indicated in fig. 3, here also repeaters are necessary at intervals in the cable. In order to be able to keep these repeaters small in spite of the large number of calls that must be amplified, the carrier waves, which would have to have a large amplitude, are themselves not transmitted (this can be accomplished by suitable connections of the modulator), but are added again in the receivers [3]).

Arrangements and functioning of a carrier-wave telephone connection

Fig. 6 is a simplified diagram of a carrier-wave connection between two places L (left) and R (right). At L are the subscribers A, C and E; at R the subscribers B, D and F. A speaks with B,

[3]) In radio broadcasting, where the available frequency region is also occupied as fully as possible, both the side bands are always transmitted, in contrast to the case in carrier-wave telephony. If in that case only one side band were transmitted, the receiver would have to add the carrier wave again. This is not practicable for a receiver which must be capable of reproducing many broadcast transmitters. If one side band plus the carrier wave were transmitted the distortion which occurs at great depth of modulation would not be permissible. In carrier-wave telephony a carrier-wave with large amplitude is added in the receiver, so that the distortion remains small. Since the microphone and the telephone already give considerable distortion the slight increase in distortion due to the lack of the second side band is not disturbing.

C with D and E with F. Between R and L lie two pairs of cores, one for transmitting speech from

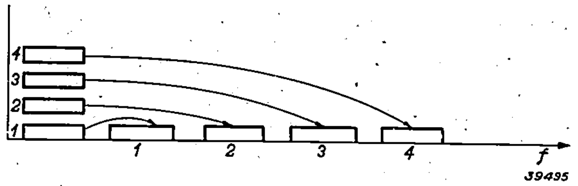


Fig. 5. Position in the frequency spectrum of the speech vibrations and the corrsponding frequency regions to be transmitted in the case of a four-channel system with suppressed carrier wave and a single side band.

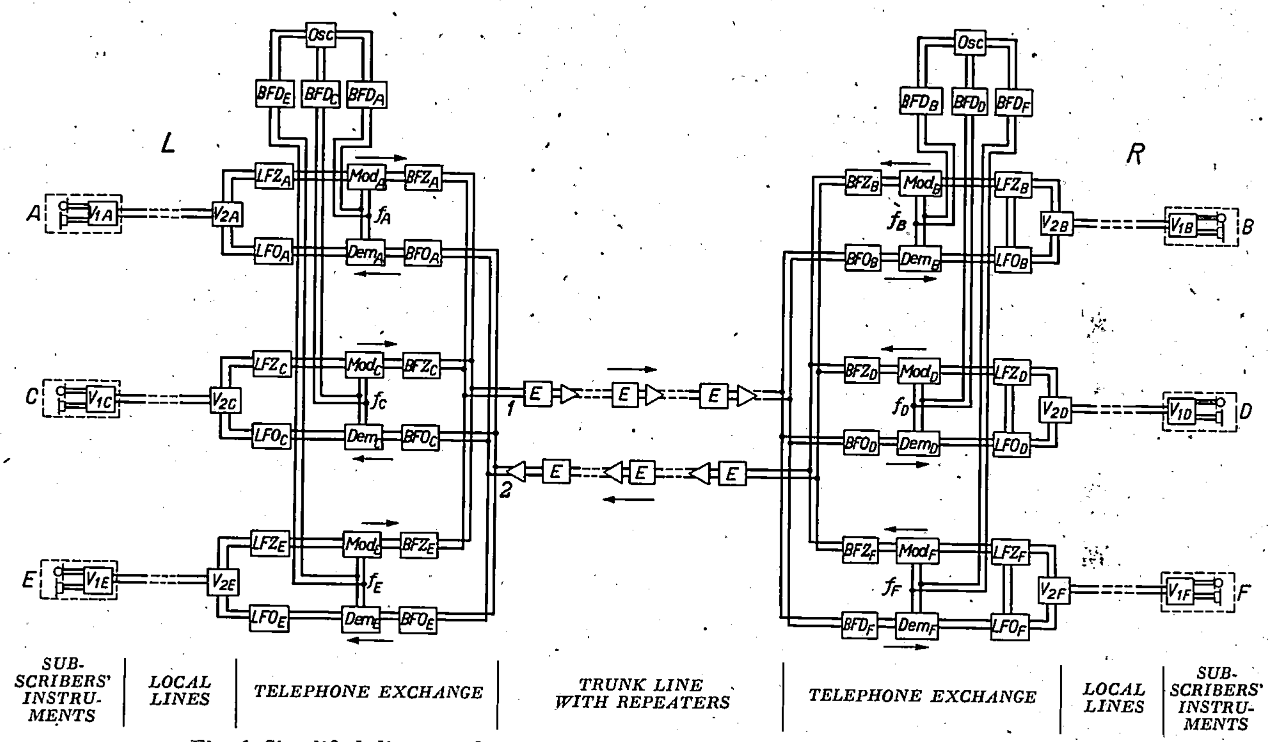


Fig. 6. Simplified diagram of a carrier-wave telephone installation for three speech channels. The subscribers' instruments consist of a microphone, a telephone and a fork V_1, and are connected by means of the local line to a fork V_2 in the telephone exchange. Each outgoing call passes a low-pass transmission filter *LFZ*, a modulator *Mod* and a transmission band filter *BFZ*. All calls are then sent together over the trunk line, which contains at intervals an equilization network E with a repeater, to the other telephone exchange, where each call passes a reception band filter *BFO*, a demodulator *Dem* and a low-pass reception filter *LFO*. By way of the fork V_2 and the local line the call reaches the other subscriber. The carrier-wave frequencies in both telephone exchanges are derived from an oscillator *Osc* by means of the carrier wave band filters *BFD*.

L to R and one for transmitting speech from R to L.

At the disposal of the subscribers are ordinary telephone instruments (microphone, telephone, fork) and they are connected to the telephone exchange by an ordinary two-wire connection. A subscriber wishing to make a call is connected, either automatically or by an operator, to an available fork V_2 in the same way as in ordinary telephony he is connected to an available line. The microphone current from A is sent through a low-pass transmitting filter LFZ_A which cuts off the frequencies above a frequency f_1 which are unnecessary for intelligibility. The frequency region which remains is fed to a modulator *ModF*, to which the carrier wave with a frequency f_A is also fed. This modulator (which functions in the same way as the modulator in a broadcast transmitter) is an arrangement which gives at its output the two side bands without the carrier wave. Following the modulator is a transmission band filter BFZ_A which passes one of the side bands but suppresses the other. The band to be transmitted may be either the upper or the lower one.

The microphone currents from C and E are treated in the same way, except that the carrier waves fed to the modulators have different frequencies, namely f_C and f_E.

The output terminals of the three transmission band filters BFZ_A, BFZ_C and BFZ_E are connected in parallel. The cable thus transmits three frequency bands: from f_A to $f_A + f_1$, from f_C to $f_C + f_1$ and from f_E to $f_E + f_1$, when the upper side bands are chosen.

At the receiving end (R) each frequency band must reach the correct channel and cause no interference in the other channels. The **reception band filters** BFO_B, BFO_D and BFO_F provide for this, each of which passes the correct frequency band but cuts off the others. Each of these band filters is followed by a demodulator to which the corresponding carrier wave is also again fed. Thus at the output terminals of the demodulator the original low-frequency band is produced, similar to that which was furnished by the microphone at the transmitting end (L), and in addition oscillators with higher frequencies. Each demodulator is followed by a **low-pass reception filter** which cuts of the undesired oscillations of high frequency. *Via* the fork V_{2B} and the local line the words of subscriber A now reach the telephone of subscriber B. In exactly the same way the words of B reach the telephone of A. In principle it is not necessary to use the same carrier-wave frequency for the conversation in the two directions between A and B; in the case of a four-wire connection as represented in fig. 6, however, it is usually done for the sake of simplicity.

The number of connection possibilities between L and R is equal to the number of pairs of double wires present multiplied by the number of calls which can be transmitted per four-wire connection. This total number is made equal to the maximum number of conversations to be carried on simultaneously between L and R.

The subscribers' instruments undergo no change in carrier-wave telephony; the subscribers observe no difference whether their call is put through with or without carrier-wave telephony.

In the example of fig. 6 three subscribers have been assumed at each end (3-channel system). It is, however, clear that this number can be extended. A 12-channel system (12 calls over the same conductor) according to this simple principle making use of ordinary, non-loaded telephone cables [1]) is often used. Philips have, moreover, designed an installation with 17 channels on this principle. *Fig. 7* is a photograph of part of the terminal apparatus necessary for this. If it is desired to transmit an appreciably larger number of calls over the same connection, a special cable is necessary and **multiple modulation** must be applied which necessitates a more complicated terminal apparatus (see below).

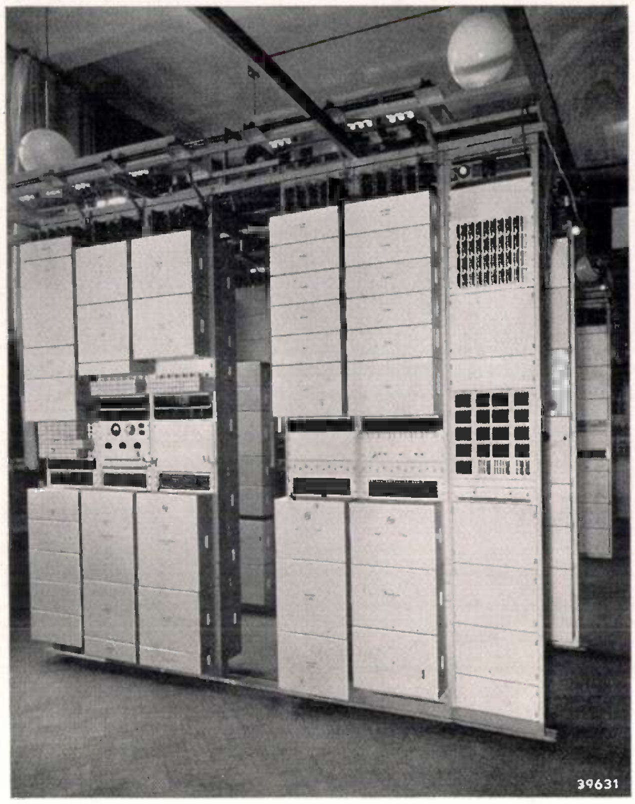


Fig. 7. Photograph of a part of the terminal apparatus of the Philips 17-channel system.

Choice of the carrier-wave frequencies

In order to make the most intense possible use of the available frequency region which the cable transmits without too great damping, the frequency difference between the carrier waves is taken as small as possible. What now determines the necessary frequency difference?

As will be seen from the foregoing, the requirement is made of each band filter that it shall pass a certain frequency region and suppress a number of other frequency regions. Between the region of transmission and the damping region there is a transitional region where the damping is too high to permit satisfactory transmission, but too low to ensure sufficient suppression. *Fig. 8* gives as an example the damping in dB as a function of the frequency for a band filter of the Philips 17-channel

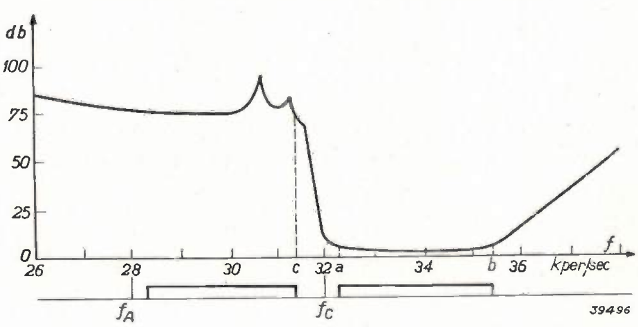


Fig. 8. Damping characteristic of a band filter. The carrier-wave frequencies of two neighbouring channels are f_A and f_C. The upper side band belonging to the carrier wave f_C fits into the transmission region of the filter (from a to b). The upper-side band belonging to the carrier wave f_A falls in the damping region of the filter (below c). The region c to a is the transitional region of the filter.

system, in which the upper side band is transmitted. The transmission region extends from a to b. To the left of it from c to a is the transitional region which has a width of about 900 c/s. (An appreciably narrower transitional region would make very much higher demands on the construction of the filter). If we assume that the curve represents the damping of the transmission band filter BFZ_C of fig. 6, then a must correspond to the lowest and b to the highest speech frequency of C to be transmitted, *i.e.* to f_2 and f_1, respectively. The highest frequency of the neighbouring channel (the speech from A to B) may not be higher than c. The distance between the two highest frequencies of the neighbouring channels, which is of course equal to the distance between the carrier-wave frequencies, must therefore be at least $f_1 - f_2 + 900$ c/s.

We have already seen that as a rule the carrier waves are not transmitted, but added to the transmitted side band again upon demodulation. It is then necessary that the carrier-wave frequencies should correspond exactly at the transmitting and receiving ends. Now in order not to require many oscillators at the receiving end as well as at the transmitting end, each of which must furnish an accurately determined frequency, one oscillator is placed at the transmitting end and one at the receiving end, whose frequencies are carefully kept constant, and the other carrier-wave frequencies are derived from these by frequency multiplication. The frequency of these oscillators is therefore chosen equal to the frequency difference between the carrier waves, namely at least $f_1 - f_2 + 900$ c/s.

As lower limit f_2 of the frequency region to be transmitted the already mentioned value of 300 c/s is generally chosen. The lower the upper limit f_1 is taken, the smaller the necessary carrier-wave distance, but the poorer the intelligibility of the speech. If necessary $f_1 = 2600$ c/s would suffice. In order, however, to be able to obtain a better quality of reproduction, and guided by the desire to have round numbers for the carrier-wave frequency (in order to be able to compare these frequencies simply with the standard frequency present in every telephone exchange, which is very precisely 1000 c/s), the choice of the carrier-wave frequencies has been determined as integral multiples of 4000 c/s [4]). From this it follows that the highest speech frequency to be transmitted may amount to $4000 - 900 + 300 = 3400$ c/s.

In the case of the installation designed by Philips for 17 calls, the lowest carrier-wave frequency, in agreement with the above, is 4000 c/s and the highest 68,000 c/s.

Cross talk

A first requirement in telephony is that each subscriber shall be able clearly to understand the speech intended for his ear, undisturbed by a background of interfering noises from conversations not intended for him. What a subscriber may hear of such noises is called cross talk. In the use of carrier-wave telephony the danger of cross talk is naturally greater than in other cases.

If the frequency spectrum of a call reaches the wrong subscriber in its original form, it is called intelligible cross talk, since with sufficient intensity the conversation can be understood. If this frequency spectrum, in its unwanted transmission, is, however, so distorted that it is impossible to follow the conversation even with sufficient intensity, it is called unintelligible cross talk. It is clear that for the sake of privacy the intensity of intelligible cross talk must be so low that it is absolutely impossible to follow the unintentionally overheard conversation. For this purpose the intensity of intelligible cross talk must be at least 70 dB below the intensity of the required transmission. Unintelligible cross talk does not violate the subscribers' privacy, but manifests itself as a disturbing background noise. An intensity about 60 dB below that of the required conversation is the limit permissible for this in practice.

It is mainly the filters which must prevent cross talk.

Requirements which must be satisfied by the parts of a carrierwave telephone installation

It may be seen from fig. 6 that the installation for carrier-wave telephony consists chiefly of cables, repeaters, oscillators, modulators and demodulators, forks and filters. Certain requirements must be made of each of these components for a satisfactory functioning of the installation. We shall consider these requirements in the following.

a) Cables

A coil-loaded cable is unsuitable for the transmission of the high frequencies necessary for a 12 or 17-channel system, due to the existence of a cut-off frequency (*fig. 9*) (see article [1]). Fig. 9 also shows the damping of an ordinary non-loaded cable in dB per km length of cable as a function of the

[4]) On the recommendation of the C.C.I.F., Oslo 1938.

frequency. The damping in the frequency region to be used (4 to 72 kc/s) increases with the frequency. Per km length of cable the difference in damping between the lowest and the highest channel (4 and 72 kc/s, respectively) is 1 dB. For

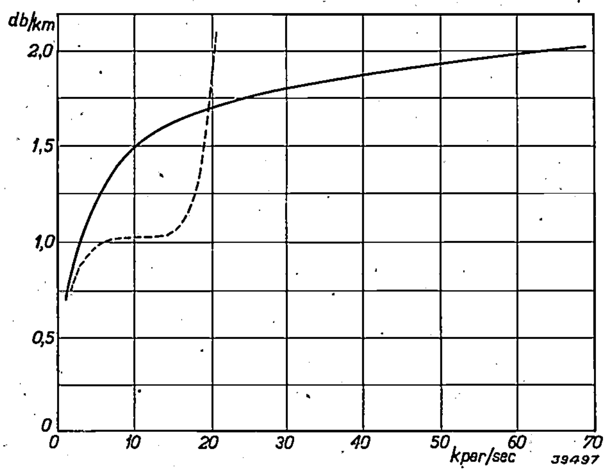


Fig. 9. Damping in dB per km as a function of the frequency for a normal, non-loaded cable. For the sake of comparison the variation of the damping is also indicated (broken line) for a lightly loaded cable.

a length of cable of 300 km (a fairly normal length for carrier telephony) this difference in damping is thus 300 dB. In connection with the combating of cross talk, however, the requirement is made that all calls shall be transmitted with about the same intensity. For this purpose so-called equalization networks are connected in series with the cable. These are quadripoles built up of resistances, self inductions and capacities in such a way that they give a high damping for the low frequencies and a low damping for the high frequencies, with the result that the damping of the cable plus the equalization network is practically the same for all frequencies considered. These requirements cannot practically be fulfilled with a single equalization network: the difference in damping at the lowest and the highest frequency is too great for that. Therefore the cables are divided into sections of such a length that it is practically possible to make the damping of each of these sections independent of the frequency with the help of one equalization network.

b) Repeaters

The damping provided by the cable with the equalization network must be eliminated again by repeaters. In the case of fairly long connections, however, it is impossible to achieve this with a single repeater either at the beginning or at the end of the line. With a cable length of 300 km and a damping of 2 dB/km a total of 600 dB of amplification is needed. If a single repeater were placed at the end of the line, then with a normal value of for instance 1 volt at the beginning, the voltage at the end of the cable would be 10^{-30} volt. But the noise voltage is of the order of 10^{-7} volt per speech channel; the speech, even after sufficient amplification, would be inaudible because of the fact that the noise voltage is 10^{23} times as high. The voltage of the signal must in fact be much higher than the noise voltage at every point in the cable, because if the ratio between signal and noise voltage has once become too small, no improvement is possible by amplification. Therefore, before the signal voltage has fallen too low due to cable damping, a repeater must be introduced. If the requirement is made that the signal-to-noise ratio must be at least 1000 at the end of the cable, it is found that a repeater must be inserted in the cable approximately every 30 km (with an amplification of about $30 \times 2 = 60$ dB).

It is a fortunate circumstance that this cable length of about 30 km is not too great to be able to make the damping independent of the frequency with a single equalization network. An equalization network is consequently placed as a rule at the end of each cable section.

The amplification may not be appreciably affected by spontaneous changes in the repeaters (decrease in the slope of the amplifier valves due to age, etc.). If for example the amplification of each repeater should decrease merely 1 dB (*i.e.* about 10 per cent), then with 10 repeaters in a line the total amplification would decrease 10 dB, which is not permissible.

Each repeater must amplify simultaneously all the calls transmitted. This makes very high demands on the linearity of the repeater. If the output voltage were not sufficiently accurately proportional to the input voltage, A.C. voltages with combination frequencies would occur, which cause mutual interference of the channels, and in addition under certain circumstances cross modulation, with the result that a subscriber would be able to hear conversation not intended for him.

By the application of inverse feed-back in the repeaters the requirements mentioned can be satisfied.

The cable damping depends upon the temperature. The amplification must therefore change in a direction opposite to the change in temperature. A temperature compensation, preferably automatic, is therefore necessary.

c) *Oscillators*

If the corresponding carrier-wave frequencies differed at the transmitting and receiving ends, the frequency spectrum of the speech would be shifted at the receiving end. This causes distortion and, with any large degree of frequency shift, unintelligibility. In practical cases in the transmission of the spoken word a frequency difference of not more than about 10 c/s between the corresponding carrier waves at transmitting and receiving end is permissible. For a carrier wave of 68 kc/s therefore a discrepancy of only about 0.015 per cent is allowed. The oscillator of 4 kc/s from which the carrier waves are derived may therefore also not vary by more than 0.015 per cent, *i.e.* about 0.05 c/s upon variations in the supply voltage, the temperature, etc.

The different carrier waves are derived from this 4 kc/s oscillator by using the successive harmonics. Each carrier wave must, however, be sinusoidal; if the carrier wave, which is fed to a given modulator or demodulator, contains a component with the frequency of one of the other carrier waves, this may lead to cross talk (intelligible as well as unintelligible). Therefore the filters which sift out the different carrier-wave frequencies (*BFD* in fig. 6) must satisfy fairly high requirements.

d) *Modulators and demodulators*

The microphone current and the carrier wave are fed to the modulator; at the output terminals the two side bands without the carrier wave are desired. In order to accomplish this the modulator contains two or more non-linear elements (cuprox or selenium rectifiers [5]) for example) in push-pull connection. An asymmetry in this connection will cause the appearance of undesired components in the output voltage. Great care must therefore be taken to secure symmetry; moreover, this symmetry must be retained in spite of temperature fluctuations and ageing.

e) *Forks*

Let us consider the closed circuit in fig. 6: V_2A-Mod_A- line 1-Dem_B-V_{2B}-Mod_B-line 2-Dem_A-V_2A. In this circuit there are, among other elements, a number of repeaters. There is therefore the danger that, following the circuit around, the total amplification will be so high that spontaneous oscillation may occur. To prevent this, care must be taken by means of accurate construction that the currents reaching the fork from the reception channel are unable to reach the transmission channel, and are fed exclusively to the local line.

f) *Filters*

The filters form one of the most important parts of the whole installation. Their function is to provide that each subscriber receives the calls intended for him in a satisfactorily intelligible form and to prevent cross talk. In order to determine the requirements which every filter must satisfy the spectral distribution is ascertained of the interference caused by the calls not intended for him which each subscriber would receive if no filters were introduced. From this the minimum damping can be found which the filters must provide for the different frequencies in question, in order that the cross talk may remain below the limits previously mentioned. The filters, for instance, which are passed on the path from *C* to *B* in fig. 6, and which together, therefore, must provide this damping, are: the low-pass transmission filter LFZ_C, the transmission band filter BFZ_C, the reception band filter BFO_B and the low-pass reception filter LFO_B. The required total damping may to a certain extent be divided among these four filters in different ways; the most economical solution will of course be chosen.

In the determination of the required total damping curve several other factors must be taken into account. The microphone furnishes not only the frequency region to be transmitted of 300 to 3400 c/s, but also higher frequencies occurring in speech. On the other hand the microphone and the telephone are less sensitive to the high frequencies than to the lower ones; moreover, the intensity of the higher frequencies in speech is less than that of the lower, the ear is also not equally sensitive to all frequencies. Taking all these facts into account the damping caused by the filters may be lower for the higher audio-frequency vibrations than is necessary for the lower.

Due to temperature variations the self-induction and capacities of which the filters are built up will change and the cut-off frequencies will therefore shift. The filter elements should therefore possess only a very small temperature coefficient.

As to the transmission region, the damping of the filters must be practically constant for all frequencies. Due to losses in the coils and condensers of which the filters are composed, the damping becomes greater for the frequencies close to the limits of the transmission region than for frequencies in the middle of this region. It is therefore

---

[5]) W. Ch. van Geel, Blocking-layer rectifiers, Philips techn. Rev. 4, 100, 1939.

necessary that these coils and condensers should be of good quality. Moreover, a transitional region of only 900 c/s, as assumed above, can only be attained when the losses in the filter elements are not too high.

Multiple modulation

In the case of the above described installation for carrier-wave telephony a frequency difference of 4000 c/s was assumed between the successive carrier waves. The transmission region of the band filters is the same for all channels. At higher carrier-wave frequencies therefore the relative width of the transmission region (*i.e.* the difference between the cut-off frequencies divided by the mean of the cut-off frequencies) becomes steadily smaller. The smaller this relative band width, the greater in practical cases is the difference between the damping at the edge and that at the middle of the transmission region, as a result of the losses in the coils and condensers (*fig. 10*). At the same time at

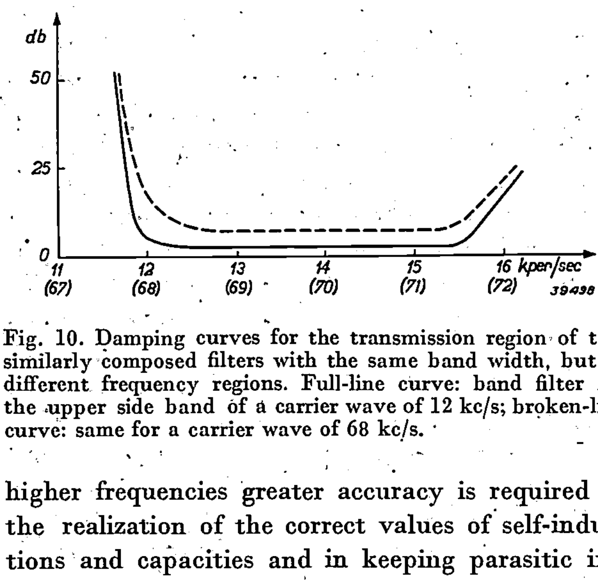


Fig. 10. Damping curves for the transmission region of two similarly composed filters with the same band width, but in different frequency regions. Full-line curve: band filter for the upper side band of a carrier wave of 12 kc/s; broken-line curve: same for a carrier wave of 68 kc/s.

higher frequencies greater accuracy is required in the realization of the correct values of self-inductions and capacities and in keeping parasitic impedances small, while the influence of temperature variations also becomes greater. For all these reasons it is in practice impossible to use carrier-wave frequencies higher than 70 kc/s in the way indicated.

Special modern cables are, however, capable of transmitting much higher frequencies (*fig. 11*). If it is desired to use this higher frequency region also for telephony, it is only possible to do so by the application of multiple modulation. This can for instance be done in the following way.

The calls to be transmitted are divided into groups of 12. Each of 12 calls is modulated in the way described above on carrier-waves with frequencies of 16 ot 60 kc/s, so that — upon the use of the upper side band — each group occupies the frequency region from 16 to 64 kc/s. Just as the frequency region of each conversation is shifted in the frequency spectrum by modulation, the frequency region of each group of twelve is now shifted

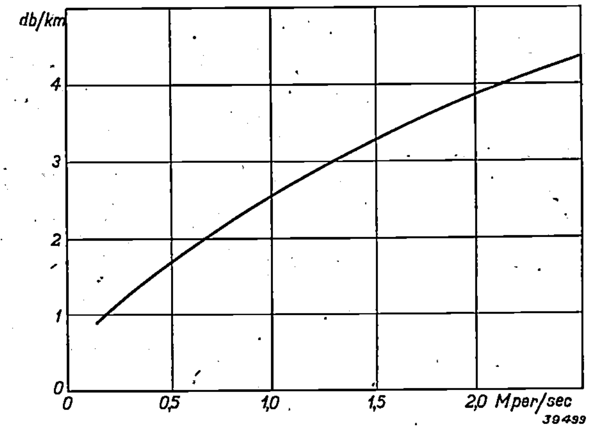


Fig. 11. Damping in dB per km as a function of the frequency for a modern coaxial cable.

by a second modulation by such an amount that the first group falls in the frequency region from 200 to 248 kc/s, the second group in the region from 248 to 296 kc/s, and so on. For 240 conversations, for example, the frequency region from 200 to 1160 kc/s is needed in this way. At the receiving end all the groups are first brought back into the region of 16 to 64 kc/s by a first demodulator, and next each conversation is shifted by a first demodulator, and then each conversation is shifted by a second demodulator into its original low-frequency region. After every second modulator a second band filter is now necessary, likewise after every first demodulator, but much less severe requirements may be made of these filters than of the first band filters, in the first place because the relative band width is high, and in the second place because the frequency difference between the side band to be passed and that to be suppressed is large so that the transitional region of the damping curve may also be wide. These filters may therefore be very simple.

In most cases, with such a large number of channels it will even be preferable to apply a third modulation for the sake of the greater simplicity of the band filters (*fig. 12*). For example, ten calls are first combined into a group which occupies the frequency region between 30 and 70 kc/s, and then five of these groups are joined to give a new group (super group) which occupies the frequency region from 300 to 500 kc/s and contains 5 × 10 = 50 calls, and finally eight of these last super groups are put side by side in the frequency spectrum by a third modulation in the region from 500 to 2100 kc/s, so that $8 \times 50 =$ 400 calls can be transmitted simultaneously.

Double modulation is also applied in the case of a small number of channels. The first modulation shifts all the calls to the same, higher, frequency region, whereupon the second modulation places

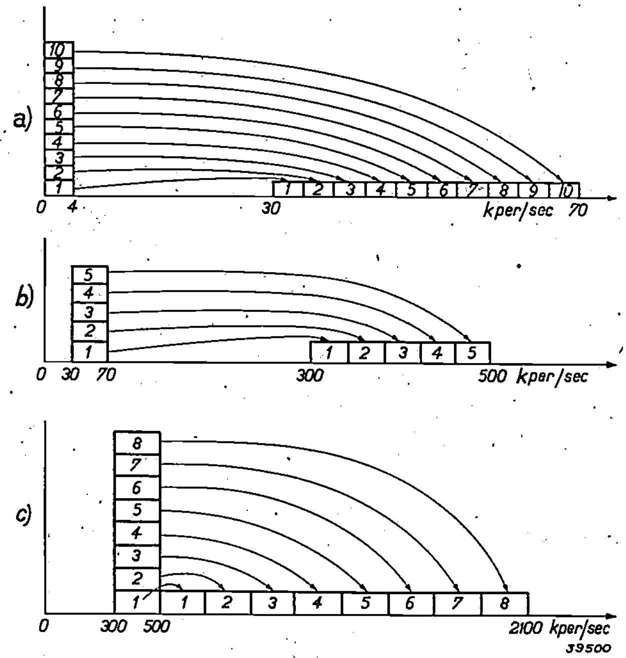


Fig. 12. Displacement of the speech vibrations in a system with triple modulation for 400 speech channels.
a Position of the frequency bands after the first modulation for each group of 10 calls.
b Position of 5 groups each of 10 calls before and after the second modulation.
c Position of 8 super groups, each of 50 calls, before and after the third modulation.

the different calls at the desired position in the frequency spectrum. After the first modulation the band filters are alike for all channels; after the second modulation they are different. The first band filters are the most complicated ones and high requirements are made of them; the second are simple filters. Since it is sometimes desirable for factory purposes that the more complicated filters should all be alike, the more complicated construction of the whole installation is sometimes preferred.

Transmission of the selector signals

Since at the present time there is a desire that not only the local connections in the telephone network but also the long distance connections between subscribers should be made without the help of a telephone operator, it is obvious that this requirement is also made of carrier-wave telephony. In ordinary telephony this is done by the subscriber making the call who sends a number of impulses to the automatic switches (selectors). Upon the use of carrier-wave telephony these impulses must first be converted in the telephone exchange into high-frequency oscillations, then transmitted by the cable and at the receiving end, after having been converted into a suitable current, they must operate the automatic selectors.

Carrier-wave telephony applied to a radio connection

In conclusion we should like to point out the possibility of transmitting different calls simultaneously with a radio connection also. To do this the currents which are sent through the cable in the carrier-wave telephony just described are modulated on the carrier wave of a transmitter (which must have a considerably higher carrier-wave frequency than the highest frequency to be transmitted), and at the receiving end this modulation is brought back again to the original position in the frequency spectrum. The four-wire connection is thus replaced by a radio transmitter and receiver at each end; the rest of the apparatus remains the same.

IMPROVEMENTS IN THE EFFICIENCY OF ELECTRIC INCANDESCENT LAMPS

by W. GEISS.

621.326

By changing over from a single coil to a coiled coil filament in the case of argon-filled electric lamps an average improvement of 15 per cent was obtained in the efficiency of the lamps. Detailed investigations were carried out on ten different kinds of lamps specially made for the purpose on the effect of krypton instead of argon as a filling. From these investigations and from results of life tests of krypton-filled coiled coil electric lamps on the market it may be concluded that with the same life and luminous flux an improvement of about 3 per cent in the average efficiency could be obtained by changing over from argon to krypton. In connection with the cost of the krypton filling this improvement cannot be realized at once in practice.

Introduction

The steadily increasing need of better and more efficient illumination can be satisfied more successfully, the more efficient the production of artificial light. With the conversion of electrical energy into visible radiation in or by means of a gas discharge the large lamp factories have followed fundamentally new methods and have developed extremely efficient sources of light which have already found an important and partially even quite new domain of application. In addition to this important work, however, the laboratories of the industries in question have by no means neglected the further development and improvement of the electric incandescent lamp. During the last ten years various possibilities were carefully examined theoretically as well as experimentally, and have to some extent already been realized in a practical form. A survey of the results achieved will be given in the following.

Importance of a gas filling

In our discussion we shall start with the original publications [1]) on the gas-filled lamp. In these it was already shown that the characteristic properties of this lamp are determined by several physical constants of the filament and of the gas used for filling the bulb. If these constants are known the efficiency (Dlm/W) of a gas-filled lamp can be very accurately calculated. It is therefore possible to determine whether an accidentally favourable result obtained in a single experiment will also be obtained in practice.

The efficiency of a tungsten filament lamp, filled with gas and used in such a way that it will attain an average life of 1000 working hours is mainly dependent on:

1) the form of the filament,
2) the composition of the gas.

Research during the last ten years has been especially devoted to a study of this dependence and we shall here discuss the results in turn.

The form of the filament

An important possibility of improving the efficiency of gas-filled lamps, and one which has been investigated thoroughly, relates to the form of the filament itself [2]).

Langmuir [1]) showed that the heat lost by a filament depends mainly on its length and only to a very limited extent upon its diameter, and at the same time that a coiled tungsten wire does not differ from a massive wire whose diameter is the same as that of the coil, as far as the loss of heat is concerned.

Instead of Langmuir's formula for the loss of heat the following empirical formula can very well be used:

$$W = C \cdot l \cdot d^{\delta} \cdot T^{\tau}, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where
W = the loss of heat to the gas,
l = the length of the filament,
d = the diameter of the filament,
T = the absolute temperature of the filament and
C = a constant depending on the gas mixture.

The exponents δ and τ are indeed again functions of the diameter and of the temperature, respectively, but they may be considered as constants within a wide range. It is found that the experimental values are best matched by choosing:

$$\delta = 0.3.$$

A large number of measurements which have been carried out during the last 15 years have confirmed this value of δ. At the temperatures mentioned in this article of 2500 to 3000 °K, τ is about 1.6.

---

[1]) I. Langmuir, Phys. Rev., 34, 401, 1912; E. Oosterhuis, Chem. Weekbld., 14, 595, 1917.

[2]) Cf. for example: W. Geiss, Philips techn. Rev., 1, 97, 1936.

For a given gas and for the temperature T, therefore, the ratio of conduction losses for two filament's of the dimensions $l_1 d_1$ and $l_2 d_2$ is

$$W_1 : W_2 = l_1\, d_1{}^{\delta} : l_2\, d_2{}^{\delta}.$$

If by means of a different form of filament it is desired to reduce the loss of heat to one half of that of an ordinary coil, the following relation must hold:

$$2 l_2 \cdot d_2{}^{0,3} = l_1 \cdot d_1{}^{0,3}.$$

This means practically that the new filament must be slightly less than half as long as the first filament, but that it may be much thicker.

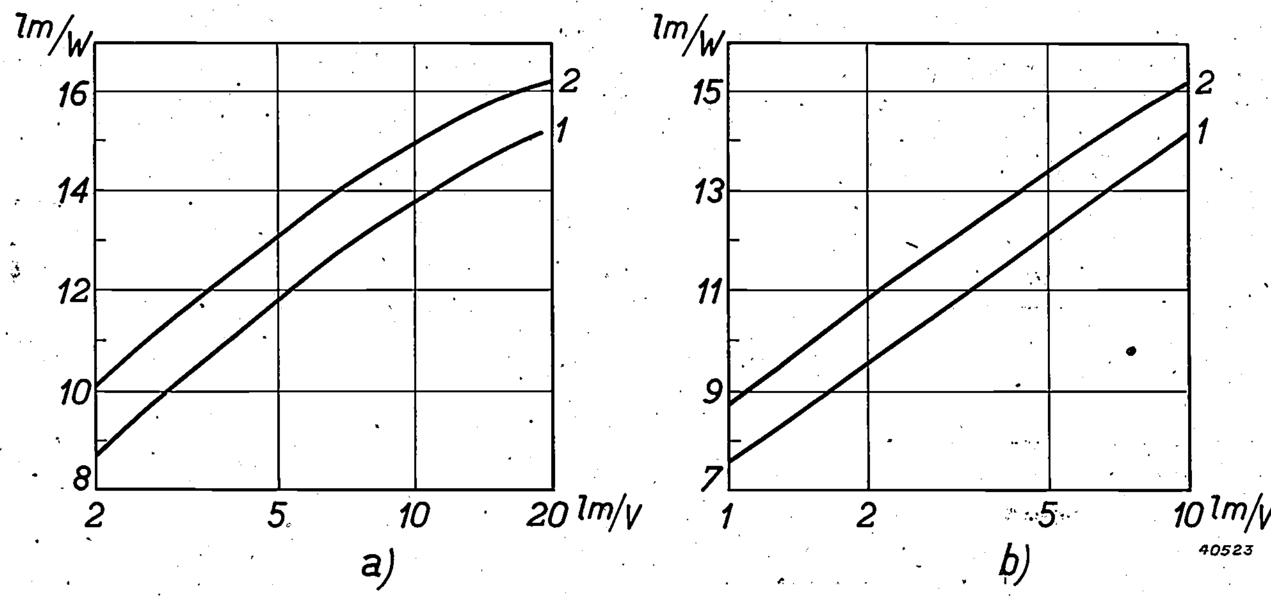


Fig. 1. The efficiency in lumens per watt for argon lamps with single coil filament (*1*) and coiled coil (*2*) as a function of the number of lumens per volt. *a*) is for lamps of 130 V and *b*) for lamps of 220—230 V.

This change in shape can be attained in two ways. A wire of the required length can be coiled on a mandrel with a diameter more than twice as great, which introduces great practical difficulties, or the ordinary coil can be coiled around a second mandrel giving the coiled coil.

In order to apply the second method, however, tungsten wire had to be available which retains its original form even at a high temperature. An elaborate investigation provided the lamp manufacturers with a solution of this problem.

On the subject of the lamp with the coiled coil, which has in the meantime been introduced everywhere and which in certain countries even occupies the foremost place, detailed articles have already appeared in this periodical [2]) and elsewhere [3]).

In *fig. 1* may be seen the curves of the efficiency for argon-filled lamps with single and coiled coil as a function of the luminous flux per volt of lamp voltage. For the commonest sizes of lamps the average improvement obtained amounts to 10.7 per cent, while for the small types it even reaches 16 per cent. This advantage for the low-power lamps is important because of the fact that upon the introduction of gas-filling in 1913 the advantage lay with the larger types.

The employment of a tungsten wire whose form does not change has also had a very favourable effect on the properties of these lamps throughout their whole life. The decrease in luminous flux compared with that of lamps with single coil has fallen to less than one half, so that the average improvement of the efficiency throughout life is still greater. According to the size of the lamps it amounts to 13—23 per cent and may be considered to average 15 per cent.

Rate of evaporation in a gas atmosphere

In addition to the form of the filament, its rate of evaporation in a gas atmosphere is very important for the efficiency of a lamp. The total radiation of a filament and the luminous flux are functions of the temperature. For their ratio E, namely the luminous flux divided by the energy radiated per second, in not too large a temperature region the following formula is sufficiently accurate:

$$E = a T^{f} \quad . \; . \; . \; . \; . \; . \; . \quad (2)$$

[3]) W. Geiss, S. E. V. Bull., **26**, 354, 1935; W. Köhler, Licht und Lampe, **24**, 462, 1935; K. Moers, Das Licht **8**, 17 and 43, 1938; K. Moers, Das Licht, **8**, 130, 1938. On the basis of the last article it can be shown that the calculation of the heat losses according to Langmuir's formula agrees well with the measured values both for the single and the coiled coils.

In the same way for the rate of evaporation v of the filament:

$$v = \gamma T^g \quad \ldots \ldots \ldots \quad (3)$$

In these expressions a, f and g are material constants of the filament, while for a given form of the filament γ depends upon the nature and the pressure of the gas surrounding the filament. This constant will be smaller the more the gas atmosphere opposes the diffusion of the vapourized molecules of the filament. Therefore, for a given filament one may speak of a diffusion constant γ of the metal in the gas with which the lamp is filled.

If one starts with the experimentally confirmed assumption that a filament has reached the end of its life when a certain part of its weight has been evaporated, its life is inversely proportional to the rate of evaporation v. With a filament having a definite life (1000 hours for instance), therefore, the rate of evaporation in different gases will be the same. Therefore the following is valid for two gases with the diffusion constants γ' and γ'':

$$v' = \gamma' \cdot T_1{}^g = v'' = \gamma'' \cdot T_2{}^g; \quad \ldots \quad (4)$$

From this it follows that:

$$\frac{\gamma'}{\gamma''} = \left(\frac{T_2}{T_1}\right)^g \quad \ldots \ldots \ldots \quad (5)$$

The following is also valid:

$$\frac{E_2}{E_1} = \left(\frac{T_2}{T_1}\right)^f \quad \ldots \ldots \ldots \quad (6)$$

By combining (5) and (6) we find that

$$\frac{E_2}{E_1} = \left(\frac{\gamma'}{\gamma''}\right)^{\frac{f}{g}} \quad \ldots \ldots \ldots \quad (7)$$

From the literature it is known [4]) that at about 2700° K the following values can be used:

$$f = 4.8 \text{ te } g = 34.3,$$

from which it follows that $f/g = 0.14$, so that

$$\left(\frac{E_2}{E_1}\right) = \left(\frac{\gamma'}{\gamma''}\right)^{0.14} \quad \ldots \ldots \ldots \quad (8)$$

From this formula it is found that a large change in the diffusion constant γ of the gas will only cause a relatively small change in E.

The diffusion constants of argon and nitrogen were determined as early as 1917 [5]), also as functions of the gas pressure, and from the measurements; moreover, it was concluded that the efficiency could be improved by using a gas with a high molecular weight. The results of the measurements were later fully confirmed by experiments by Fonda [6]) (see *fig. 2*).

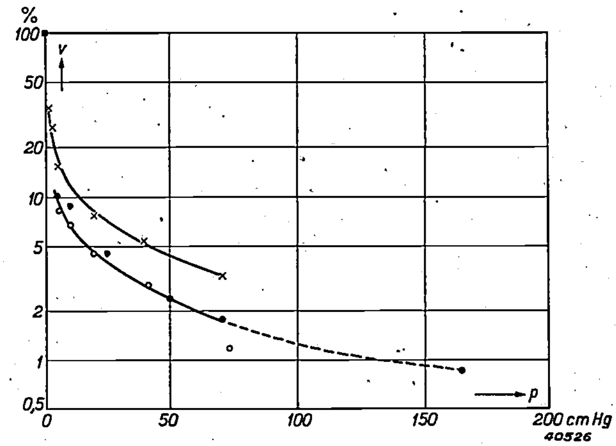


Fig. 2. Rate of evaporation v as a function of the pressure p (cm mercury), expressed in % of the rate of evaporation in a vacuum.
× nitrogen, according to Oosterhuis [5]).
○ argon with 10% nitrogen, according to Oosterhuis [5]).
● argon with 14% nitrogen, according to Fonda [6]).

In addition Fonda [7]) fixed the relation between the rate of diffusion and molecular weight in a formula, and pointed out that the calculation of the diffusion can be connected with Langmuir's theory of heat losses.

In this theory it is assumed that the cylindrical filament of thickness d is surrounded by a cylindrical film (Langmuir layer) of diameter b in which the gas does not move. According to Langmuir the following expression then holds:

$$b \ln \frac{b}{d} = k \frac{r}{s}, \quad \ldots \ldots \ldots \quad (9)$$

where k represents a constant, while r is the viscosity and s the density of the gas. If, further, p is the pressure and m the molecular weight of the gas, then according to Fonda the following formula is found for the diffusion constant:

$$\gamma = \frac{C^{\text{te}}}{m\,d\,p \ln \frac{b}{d}}, \quad \ldots \ldots \quad (10)$$

which we shall make use of in the following.

Comparison of the rates of evaporation in argon and krypton

According to (10) for two kinds of gas at the same

[4]) H. A. Jones and I. Langmuir, Gen. El. Rev., **30**, 310, 1927.
[5]) E. Oosterhuis, loc. cit.
[6]) G. R. Fonda, Phys. Rev., **31**, 260, 1928.
[7]) G. R. Fonda, Phys. Rev., **21**, 343, 1923.

pressure the ratio of the diffusion constants γ' and γ'' is the following:

$$\gamma'' : \gamma' = \frac{\ln (b'/d)}{\ln (b''/d)} \cdot \frac{m'}{m''} \quad . \; . \; . \quad (11)$$

From this the ratio of the diffusion constants γ of two rare gases, argon and krypton for instance, can be calculated.

The comparison between these two rare gases has become of real importance since the liquid gas industry has succeeded in obtaining krypton on a fairly large scale by a new process [8]) which we shall not go into here. Krypton occurs in only extremely small quantities in the air, namely only one millionth by volume.

The molecular weights of argon and krypton are

$$m_{\mathrm{Ar}} = 40 \text{ and } m_{\mathrm{Kr}} = 83.$$

The viscosities at 20° C are

$$r_{\mathrm{Ar}} = 2.21 \cdot 10^{-6} \text{ and } r_{\mathrm{Kr}} = 2.48 \cdot 10^{-6}.$$

The densities of the gases are

$$s_{\mathrm{Ar}} = 1.78 \text{ and } s_{\mathrm{Kr}} = 3.71.$$

As diameter d of the filament let us choose that of the ordinary 40 Dlm lamp, while it may be remarked that the result of the calculation scarcely changes if one takes for d the diameter of the filament of a 150 Dlm lamp. Calculating b/d of equation (9) and filling this in (11) one obtains

$$\gamma_{\mathrm{Kr}} = 0{,}59 \; \gamma_{\mathrm{Ar}} \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

The diffusion constant γ for pure krypton thus amounts to about 60 per cent of that for pure argon. This figure has only a theoretical significance, however, since experience has shown that a filling of pure rare gas cannot be used in lamps for general service. The arcing voltage of rare gases is relatively low, so that in a lamp filled with a pure rare gas an arc may occur between the electrodes which renders the lamp useless. The chance of arcing in the rare gas is considerably diminished when nitrogen is added.

The choice of the proportion of nitrogen will in general be based upon a compromise. Much nitrogen means a decrease in efficiency, little nitrogen increases the chance of arcing.

The most satisfactory nitrogen content cannot therefore be determined on the basis of a limited number of laboratory tests. It is much better to rely upon the experience of many years and the extensive statistical data furnished by the regular lamp testing on the basis of life tests.

In this way in the Philips concern experience has shown that with krypton a content of nitrogen at least 5 per cent higher than with argon must be added, if the same security against arcing of the gas is to be obtained. The arcing voltage of krypton is lower than that of argon.

In order not to make the discussion unnecessarily complicated we shall in the future, when argon and krypton are compared, calculate with the same percentage of nitrogen, namely a content of about 13 per cent which is customary for argon.

By the addition of nitrogen to rare gases the diffusion constant is affected. Fifteen years ago we investigated the rate of evaporation of tungsten wires in different mixtures of argon and nitrogen. It was then found that the rate of evaporation varies proportionally with the nitrogen content (see *fig. 3*, in which the rate of evaporation determined for krypton and pure xenon are also indicated). From this it follows that it is possible in general to calculate the diffusion constant γ of a gas mixture in a simple way from the ratio of the gases of which the mixture is composed.

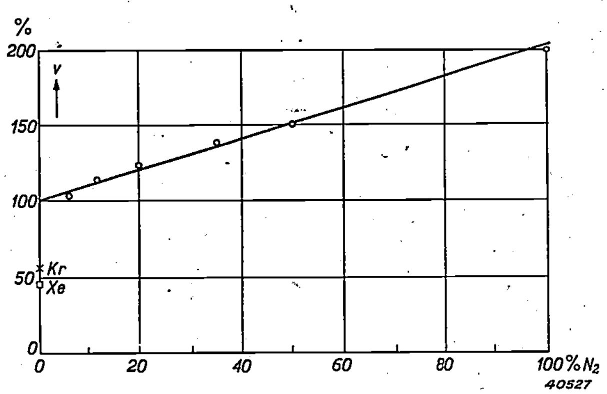


Fig. 3. Velocity of evaporation v in mixtures of argon and nitrogen as a function of the nitrogen content, expressed in % of the velocity of evaporation in pure argon. For the sake of comparison the velocities of evaporation in pure krypton × and xenon ⊡ are also indicated.

On this basis we have calculated the diffusion constants γ according to formula (10) for the ordinary lamps of 25, 40, 65, 100 and 150 Dlm with coiled-coil filament, not only of the series for 130 volts but also of that for 220—230 volts, and we have used the values found in equation (8). We then obtained the ratio of the values of E with a krypton or an argon filling. The E-values here considered are the ratios of the luminous flux in lumens to the energy radiated in watts. In order to be able to compare the results of our calculation with the results of practical measurements, where

[8]) Ph. Siedler, Angew. Chem., 51, 799, 1938.

the ratio of the luminous flux to the power consumed, the so-called efficiency, is always determined, it is important to recalculate all results on this basis. We began at first with the assumption that the heat losses of the filament to the gas filling are the same for the krypton mixture as for the argon mixture. In *fig. 4* the results of the measurements are shown, *a*) for lamps of low voltage and *b*) for lamps of high voltage. It may be concluded that for fillings of krypton and argon with the same percentage of nitrogen, the krypton will give an efficiency 4 per cent higher than that of argon, merely as a result of the difference in rate of evaporation for the same filament.

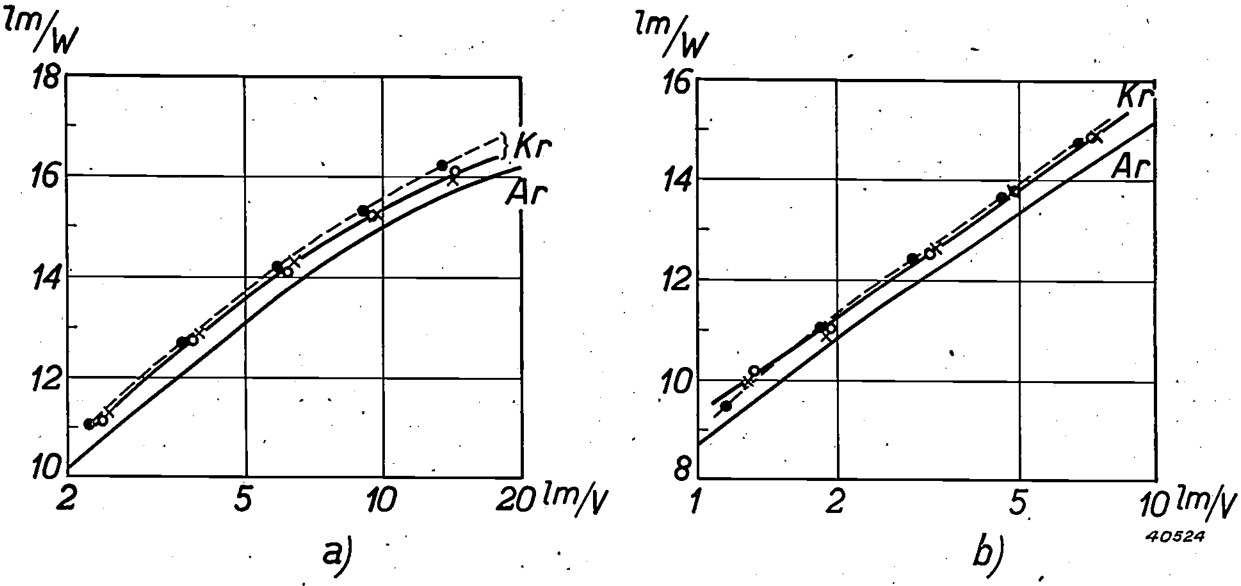


Fig. 4. The efficiency in lumens per watt is plotted as a function of the number of lumens per volt for *a*) lamps of 130 V and *b*) lamps of 220—230 V. The curves *Ar* are for argon filling and *Kr* for krypton filling, and the values calculated with formulae (10) and (8) are indicated by ●, while ○ and × follow from measurements of the luminous flux and the resistance, respectively.

Investigation of coiled-coil lamps with different gas fillings

In order to test the theoretical results of the foregoing section about 50 of each of the 10 types of lamps mentioned were specially made with great care, one half filled with a krypton-nitrogen mixture and the other half with an argon-nitrogen mixture. After having been measured these lamps were subjected to a life test.

From the difference in luminous flux upon transition from argon to krypton the difference in temperature can immediately be calculated, and from that the change in efficiency. These latter figures can also be calculated from the difference in the resistance of the coil. All results are recalculated on the basis of the same life of 1000 hours, while it is assumed that the heat losses to the gas are the same for krypton as for argon.

The values obtained experimentally by both these methods are also included in fig. 4. The points indicated by small circles are obtained on the basis of light flux measurements, those indicated by crosses on the basis of determinations of the resistance of the incandescent filament. The two series of points show good agreement, not only with each other, but also with the points calculated according to formula (10). The theoretically found improvement of the efficiency upon transition from argon to krypton is about 4 per cent, while the experiments gave the result 3.2 per cent. The difference between these two results, while small, may perhaps be explained by the fact that the krypton-nitrogen mixture used already showed some tendency towards arcing and the formation of an arc, which makes the average life of the lamps shorter.

The value of about 4 per cent according to the calculation may therefore be considered as the maximum improvement, upon which one might count only if there were no danger of breakdown of the gas and if the heat losses to the gas were the same for krypton and argon. The somewhat lower value of 3 per cent found practically indicates that there is actually a danger of arcing. In order to eliminate this danger the nitrogen percentage of the krypton mixture should be chosen slightly higher, which would result in a decrease in the improvement to be expected to about 3 per cent.

Thermodiffusion

Until now in our discussion we have taken into account only the ordinary diffusion of the evaporated tungsten through the gas. In addition to this there is the so-called "thermodiffusion" [9]) which consists in the fact that in the presence of a temperature gradient in a gas mixture the heavier gas diffuses toward the colder zone. This diffusion is more rapid the steeper the temperature gradient and the greater the difference between the densities of the gases.

Since there is a very steep temperature gradient in the Langmuir film a considerable thermodiffusion of the heavy atoms of the vapourized tungsten may be expected; this diffusion will be more rapid in the lighter argon than in the heavier krypton. The rate of diffusion of tungsten in krypton should therefore be less than the 59 per cent of that in argon indicated by formula (13).

Since the rare gases used are not pure, but are mixed with nitrogen, the influence of the nitrogen on the thermodiffusion must still be ascertained. The latter gas in the mixture is also subject to thermodiffusion. As the lighter component it will diffuse toward the filament, so that there the gas mixture will be richer in nitrogen. Around the filament, therefore, the average density of the gas atmosphere is less than that of the gas mixture used for filling. The addition of nitrogen to the rare gas increases the diffusion and thus has an effect opposite to that which results from the transition from argon to krypton.

From the experiments it may be concluded that it is unnecessary to introduce a correction for the thermodiffusion into formula (10) since the experimental results do not lie above, but always below the results of the calculation.

Colour difference between argon and krypton lamps

From equation (5), since the ratio of the diffusion constants and the exponent g are known, the increase in temperature can be calculated upon the transition from the argon to the krypton mixture. For the ten types of lamps specially made for this investigation it amounts on an average to

$$\Delta T = 0.0125\ T = 34\ ^\circ\mathrm{K}\ \text{(calculated)}.$$

From the optical and electrical data we found for these types of lamps the increase in temperature given in *table I.*

Table I

	ΔT, calculated from the increase in the light flux	ΔT, calculated from the increase of the resistance
	38 °K	45 °K
	14	24
	32	28
	28	32
	31	31
	29	24
	29	21
	28	20
	19	19
	15	21
average	26.3 °K	26.5 °K

From this table it is evident that no systematic variation of ΔT with the luminous flux of the lamps investigated could be derived. The mean error is about 7°. The increase in the true temperature of a tungsten filament upon the transition from an argon to a krypton mixture as filling for the bulb, therefore, is an average of

$$\Delta T = 0.0098\ T = 26^\circ \pm 7^\circ.$$

An investigation of krypton and argon lamps, as put on the market, with the help of the pyrometer gave a slightly lower average value for the difference, namely 20°, which may be explained by the somewhat higher nitrogen content in krypton lamps.

It need hardly be stated that it is impossible to perceive a temperature difference of 1% in electric lamps without optical instruments. This can easily be verified by a simple and yet technically convincing test. A krypton and an argon lamp are screened from each other and each used to illuminate part of a white surface. It is impossible to determine a colour difference with any degree of confidence.

Efficiency and heat losses

Since a considerable part of the energy which is taken by an electric lamp with gas filling is lost by direct heat conduction to the gas, we shall examine the influence of this on the efficiency somewhat more carefully. In the case of the 40 Dlm lamp for 220—230 volts, which is the most commonly used gas-filled lamp with single coil, almost 33 per cent of the energy was lost in this way, since it did not contribute to the light radiation. One of the means

[9]) C. Ludwig, Wien. Ber., 20, 539, 1856; Ch. Soret, Arch. Genève, 2, 48, 1879; Ann. Chim. et Phys., 22, 293, 1881; Chapman, Phil. Mag., 7, 1, 1929; Ibbs, Proc. Roy. Soc., A 99, 385, 1921.

of decreasing this loss and thus increasing the efficiency consists in using gases for filling, which have a poor heat conductivity.

According to Langmuir [1]) the heat W which a cylindrical filament gives off at the temperature T per cm length to the surrounding gaseous atmosphere with a temperature of 300 °K, is determined by the equation

$$W = \sigma\, (\varphi_T - \varphi_{300}), \quad \dots \quad (13)$$

where

$$\sigma = \frac{2\pi}{\ln\left(\frac{b}{d}\right)} \quad \dots \quad (14)$$

and $$\varphi = 4.19 \cdot \int_0^T \lambda \mathrm{d}T, \quad \dots \quad (15)$$

while b = the diameter of the Langmuir film,
d = the diameter of the filament,
λ = the heat conductivity of the gas.

The heat conductivity λ could be calculated with the help of Sutherland's formula for the viscosity and from the specific heat at constant volume. The constants occurring were taken from the Landolt-Börnstein tables.

The calculation of $(\varphi_T - \varphi_{300})$ for different temperatures T of the filament gave the values shown in *table II* for nitrogen, argon and krypton.

Table II

T	$\varphi_T - \varphi_{300}$		
	N_2	Ar	Kr
1000 °K	0.293	0.216	0.124
1500	0.623	0.453	0.264
2000	1.030	0.743	0.438
2500	1.506	1.083	0.642
3000	2.047	1.451	0.868

When these values are represented in a graph with a double logarithmic scale (see *fig. 5*) they lie on a straight line for each gas, at least the values for 2000°, 2500° and 3000° K in which we are interested, so that we may represent $(\varphi_T - \varphi_{300})$ in this temperature region which is important for electric lamps by the formula:

$$\varphi_T - \varphi_{300} = a_1 \cdot T^{\beta} \quad \dots \quad (16)$$

or also

$$\varphi_T - \varphi_{300} = a \left(\frac{T}{2\,700}\right)^{\beta}, \quad \dots \quad (17)$$

where we assume that 2700 °K is approximately the average temperature of the tungsten filament.

The constants a and β then take on the values in *table III*.

Table III

	a	β
Nitrogen	1.70	1.65
Argon	1.22	1.65
Krypton	0.73	1.65

In order to calculate the losses by heat conduction by the gas, in addition to $(\varphi_T - \varphi_{300})$ the

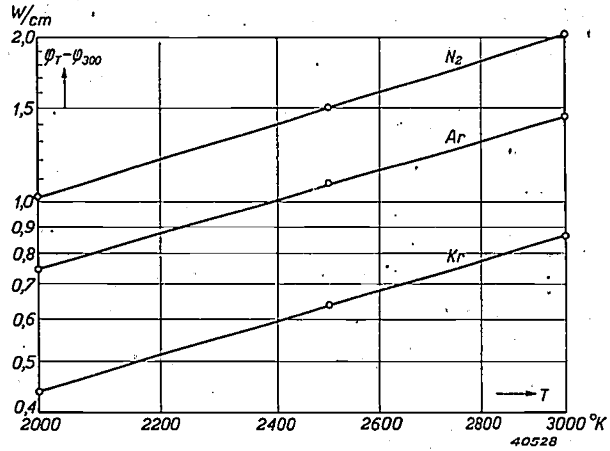


Fig. 5. The quantity $\varphi_T - \varphi_{300}$ in watts per cm as a function of the absolute temperature T for nitrogen, argon and krypton.

"shape factor" σ, which occurs in (13) must also be known. Langmuir gives the following formula for this:

$$\frac{d}{B} = \frac{\sigma}{\pi} \cdot e^{-2\pi/\sigma}, \quad \dots \quad (18)$$

where d again represents the diameter of the filament and B is the thickness of the Langmuir film for a plane surface. This formula corresponds to (9) with

$$k = \frac{s}{r}\, 2\, B \quad \dots \quad (19)$$

For air with a pressure of 76 cm of mercury, according to Langmuir's measurements,

$$B = 0.43 \text{ cm}.$$

Langmuir has further assumed that B is proportional to the ratio of the viscosity and the density, so that from the known data B can be calculated for the different gases.

For a pressure of 57 cm one then finds that:

for nitrogen B_N = 0.57 cm,
for argon B_{Ar} = 0.51 cm and
for krypton B_{Kr} = 0.27 cm.

From equation (18) it is now possible to calculate σ and thus also $W = \sigma\,(\varphi_T - \varphi_{300})$.

For $T = 2700$ °K one then obtains the heat losses given in *table IV* in W/cm for different diameters of the filament with rare gas mixtures containing 13 per cent of nitrogen.

Table IV

Diameter of the filament in μ	W/cm (2700 °K, 57 cm of mercury)		
	N (pure)	$Ar_{87}N_{13}$	$Kr_{87}N_{13}$
100	3·08	2·37	1·78
200	3·62	2·80	2·12
300	4·01	2·11	2·38
400	4·35	3·38	2·61
500	4·66	3·62	2·82
700	5·18	4·03	3·20
1000	5·85	4·58	3·69

If the dimensions of the filament are known the energy lost by heat conduction can be calculated for the different gas mixtures, and on the basis of that, the way in which the efficiency is changed can be determined. This calculation was carried out for the ten specially made types of lamps; the results, for an argon mixture as well as for a krypton mixture are given in *fig. 6.*

It may be seen that when the argon mixture is replaced by a krypton mixture the efficiency in Dlm/W increases on an average of 3.6 per cent due to the smaller heat losses, when we assume that the rate of evaporation of tungsten is the same in argon as in krypton. This value will, however, fall to 2-3 per cent, when, taking into account the greater chance of arcing in krypton, more nitrogen is added than is customary in the case of argon.

Final considerations

In this article we have studied successively the influence of the form of the filament and of the composition of the gas filling on the efficiency of lamps, the average life remaining the same. By passing over from single coil to coiled-coil filaments of the same luminous flux and life, an improvement of an average of 15 per cent in the efficiency was obtained. With this, however, the possibilities of development of the gas-filled lamp are by no means exhausted. Since the introduction of lamps with coiled-coil filaments a further improvement of the efficiency of this type has been obtained on the basis of earlier fundamental research [10]), which, however, we shall not go into here.

According to the discussion given in this article it might be expected that by making use of krypton instead of argon for filling coiled-coil lamps a further improvement could be obtained as a result of the differences in rate of evaporation and in heat conduction. It must, however, be kept in mind that the improvement relates to the initial value of the efficiency. Since, however, during use this last quantity decreases more rapidly with krypton lamps than with argon lamps, in practice only a slight improvement will be obtained. Especially in large types of lamps this phenomenon occurs, be-

[10]) E. Oosterhuis, loc. cit.

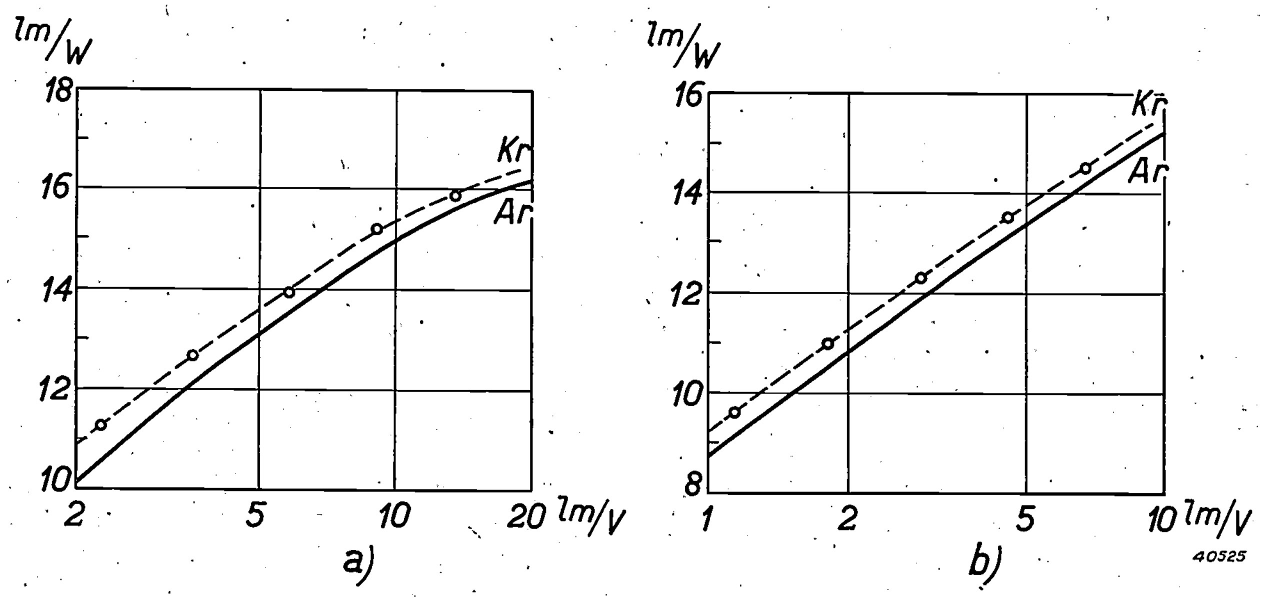


Fig. 6. The efficiency in lumens per watt which is obtained taking into account the different energy losses by heat conduction from the filament is plotted as a function of the number of lumens per volt. The curves are for argon (*Ar*) and krypton (*Kr*) mixed with 13% of nitrogen to prevent breakdown. *a*) is for lamps of 130 V and *b*) for lamps of 220—230 V.

cause due to the high cost of krypton relatively small bulbs must be used for krypton lamps, so that they become blacker than is the case with the corresponding argon lamps. In agreement with this we found that the lamps on the market with krypton filling and coiled coil, compared with our corresponding coiled-coil argon lamps of the same luminous flux, exhibited an improvement of 3 per cent in the efficiency throughout life. To what degree it is in general justifiable to realize this improvement, considering the cost of the krypton filling, is an economic problem which we shall not discuss here.

In general that development will be preferable by which a technical advance can be attained without an intolerable increase in production costs, and with which, moreover, the new lamp can replace the existing type almost completely. A further requirement of the general utility of a new technical improvement is that it shall not have a retarding action on later possibilities of development. *The coiled-coil lamp filled with argon* and in normal dimensions satisfies these conditions. In recent years it has gained an absolutely dominating position with the result that the former lamp with single coil has practically become obsolete while it leaves the path open for further improvements in the future.

ON THE CONSTRUCTION OF VIBRATORS FOR RADIO SETS

by J. KUPERUS. 621.314.5 : 621.396.02

Several problems are discussed, which are connected with the construction of vibrators for the connection of an A.C. receiving set to the D.C. mains. The study of these problems has led to the development of a new type of vibrator which is described.

A description has already been given in this periodical of a vibrator which serves for the connection of A.C. receiving sets to a D.C. main [1]). In this article we shall discuss a number of problems connected with the construction of such vibrators and describe a new type of vibrator in which various improvements are incorporated.

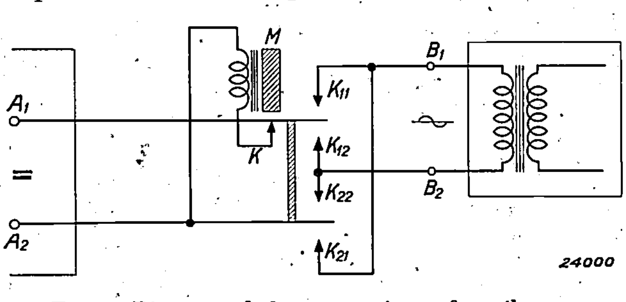


Fig. 1. Diagram of the connections of a vibrator.

The connections of a vibrator are represented diagrammatically in *fig. 1*. The springs A_1 and A_2, which are insulated from each other electrically but joined mechanically, can alternately make contact against K_{11} and K_{12}, and K_{22} and K_{21}, respectively. The springs are moved by the electromagnet M acting on an armature which is mechanically connected with the springs A_1 and A_2. When these springs are in the stationary state the spring A_1 is connected with the coil of the magnet by means of the contact K. When the vibrator is connected to the D.C. mains the electromagnet will attract the armature and contact will be made between A_1 and B_1 and between A_2 and B_2. At the same time the contact K is broken so that the armature is no longer attracted. Due to the effect of inertia the springs continue to move still farther and then swing back through the stationary state, making contact between A_1 and B_2, and A_2 and B_1, respectively. Due to the fact that the magnet is again excited each time in a suitable phase of the vibration of the springs, the mechanism keeps itself going. A current thus flows through the primary winding of the transformer which continually changes its direction, so that in the secondary winding an A.C. voltage is induced whose magnitude depends upon the voltage of the D.C. mains, among other factors, and on the ratio of the number of windings of primary and secondary coils of the transformer.

Due to the self-induction of the transformer and the magnet coil, with connections like those of fig. 1, a high voltage would occur between the points of

[1]) Philips techn. Rev., 2, 346, 1937.

cause due to the high cost of krypton relatively small bulbs must be used for krypton lamps, so that they become blacker than is the case with the corresponding argon lamps. In agreement with this we found that the lamps on the market with krypton filling and coiled coil, compared with our corresponding coiled-coil argon lamps of the same luminous flux, exhibited an improvement of 3 per cent in the efficiency throughout life. To what degree it is in general justifiable to realize this improvement, considering the cost of the krypton filling, is an economic problem which we shall not discuss here.

In general that development will be preferable by which a technical advance can be attained without an intolerable increase in production costs, and with which, moreover, the new lamp can replace the existing type almost completely. A further requirement of the general utility of a new technical improvement is that it shall not have a retarding action on later possibilities of development. *The coiled-coil lamp filled with argon* and in normal dimensions satisfies these conditions. In recent years it has gained an absolutely dominating position with the result that the former lamp with single coil has practically become obsolete while it leaves the path open for further improvements in the future.

ON THE CONSTRUCTION OF VIBRATORS FOR RADIO SETS

by J. KUPERUS.

621.314.5 : 621.396.02

Several problems are discussed, which are connected with the construction of vibrators for the connection of an A.C. receiving set to the D.C. mains. The study of these problems has led to the development of a new type of vibrator which is described.

A description has already been given in this periodical of a vibrator which serves for the connection of A.C. receiving sets to a D.C. main [1]). In this article we shall discuss a number of problems connected with the construction of such vibrators and describe a new type of vibrator in which various improvements are incorporated.

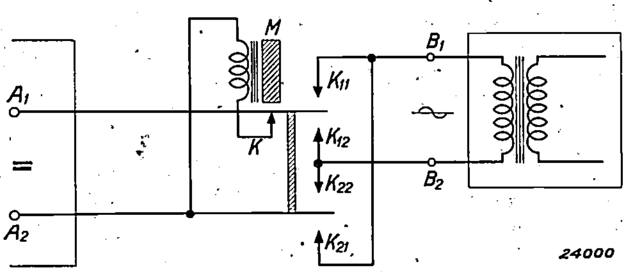


Fig. 1. Diagram of the connections of a vibrator.

The connections of a vibrator are represented diagrammatically in *fig. 1*. The springs A_1 and A_2, which are insulated from each other electrically but joined mechanically, can alternately make contact against K_{11} and K_{12}, and K_{22} and K_{21}, respectively. The springs are moved by the electromagnet M acting on an armature which is mechanically connected with the springs A_1 and A_2. When these springs are in the stationary state the spring A_1 is connected with the coil of the magnet by means of the contact K. When the vibrator is connected to the D.C. mains the electromagnet will attract the armature and contact will be made between A_1 and B_1 and between A_2 and B_2. At the same time the contact K is broken so that the armature is no longer attracted. Due to the effect of inertia the springs continue to move still farther and then swing back through the stationary state, making contact between A_1 and B_2, and A_2 and B_1, respectively. Due to the fact that the magnet is again excited each time in a suitable phase of the vibration of the springs, the mechanism keeps itself going. A current thus flows through the primary winding of the transformer which continually changes its direction, so that in the secondary winding an A.C. voltage is induced whose magnitude depends upon the voltage of the D.C. mains, among other factors, and on the ratio of the number of windings of primary and secondary coils of the transformer.

Due to the self-induction of the transformer and the magnet coil, with connections like those of fig. 1, a high voltage would occur between the points of

[1]) Philips techn. Rev., 2, 346, 1937.

contact each time a contact was broken. The sparks thus caused would considerably diminish the life of the contacts. Because of this the contacts are shunted with condensers.

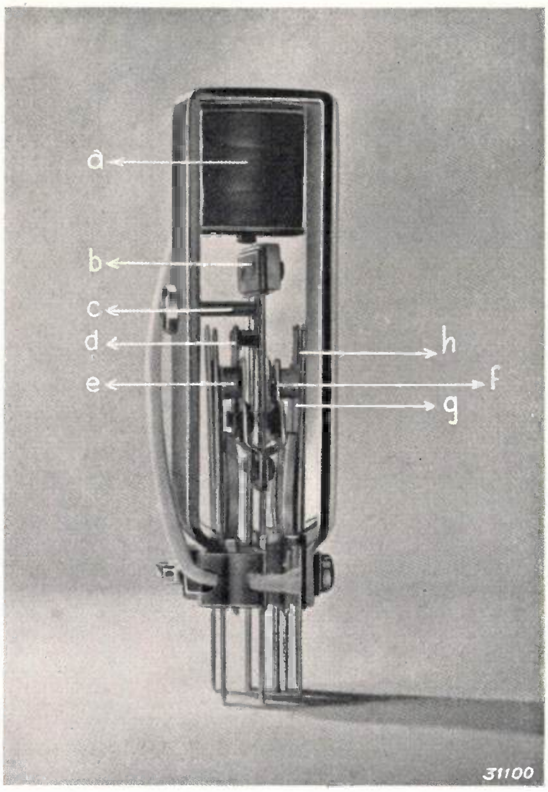


Fig. 2. New construction of the vibrator.
a = magnet coil,
b = armature,
c = stop for the spring *D* (see fig. 5),
d = contact for switching the current into the magnet coil,
e = contact for changing the direction of the current in the external circuit,
f, *g* and *h* may be compared with *1*, *2* and *3* in fig. 7.

In order to prevent the penetration into the mains or into the receiving set of the high-frequency oscillations which occur due to the sudden current variations in the vibrator, the connection to the mains as well as to the receiving set is *via* filters. The complete vibrator thus consists of two parts, namely:

1. the vibrator proper
2. the coil and condensers for the filters und shunts across the contacts, built together as a compact unit. This part will be called the anti-interference part.

Both parts are housed together in an acoustically and electrically shielding container.

Fig. 2 shows a new construction of the vibrator. We shall now discuss several problems whose solution led to the construction illustrated in this figure.

The attachment of the vibrator to the anti-interference part

Since the vibrator contains moving parts, if it were rigidly fastened to the other parts the latter, including the container, would also be set vibrating. The result would be that sound vibrations would be transmitted to the surrounding air. In order to prevent this, the attachment of the vibrator to the anti-interference part is made in such a way that no vibration is transmitted to the anti-interference part upon vibration of the springs. In order to illustrate this method of fastening, the centres of gravity of certain parts of the vibrator are indicated in *fig. 3a*. *A* is the centre of gravity of the whole vibrator. Since with the method of fastening to be described no external forces act upon the vibrator due to the motion executed by the latter, the centre of gravity *A* will not change its position during this motion. *B* indicates the centre of gravity of the moving parts (armature and springs) while *C* represents the centre of gravity of the remainder of the vibrator. This remainder consists of the electromagnet with coil (centre of gravity *D*), the frame and the attachment of the springs (centre of gravity *E*). If *B* moves in a certain direction as a result of a deviation of the armature and the springs, then, since *A* remains stationary, *C* will move in the opposite direction. In fig. 3*b* this motion of the centres of gravity is represented. Since the part whose centre of gravity is formed by *B* does not execute a pure translation, but a movement which may in general be described as a translation

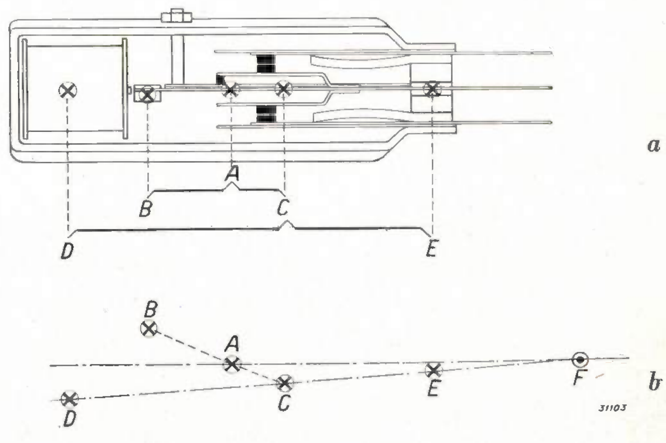


Fig. 3. *a*) Position of the centres of gravity of different parts of the vibrator.
b) Relative displacement of these centres of gravity.
A = centre of gravity of the whole,
B = centre of gravity of armature + springs;
C = centre of gravity of *D* + *E*; where
D = centre of gravity of electromagnet with coil,
E = centre of gravity of frame + attachment of springs,
F = point of attachment of the vibrator to the anti-interference part.

of the centre of gravity B and a rotation around the centre of gravity, the motion of the rest will also possess this general character, in which the rotation as well as the translation is in the opposite direction. This combined movement of the rest can be described as pure rotation around the point F in fig. 3b. This point therefore does not change its position in the movement executed by the vibrator. In the new construction of the vibrator the attachment of the vibrator to the anti-interference part takes place exclusively in the neighbourhood of point F, so that the motion of the vibrator can no longer be transmitted to the rest of the apparatus.

Since this fastening is flexible, upon fairly rough treatment in the transport of the vibrator it may easily be damaged. To prevent this the vibrator proper is so constructed that it can be fastened to the anti-interference part by means of a base which shows some similarity with a valve holder. The vibrator can therefore be taken out of the apparatus for transport and packed separately.

Improved construction of the contact for sending the current through the magnet coil

In older types of vibrators the contact K is always so constructed that the stop for the spring A_1 (see fig. 1) also serves as contact point. While this gives simplicity of construction, the objections involved in this system have, nevertheless, led to an alteration.

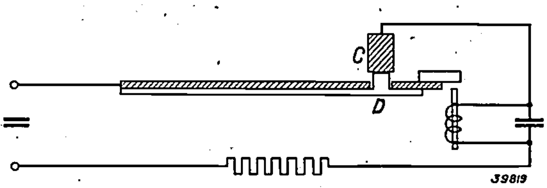


Fig. 4. Construction of the contact for sending the current through the magnet coil. The stop C of the spring D forms at the same time the contact.

In *fig. 4* the older construction here referred to is shown diagrammatically. The fixed stop C thus serves at the same time as contact point, and together with D forms the contact K of fig. 1. A clear insight into the objection to this construction is obtained by a comparison with a hammer which strikes a heavy anvil. After the first contact with the anvil the hammer will rebound and again fall on the anvil, and the process will be repeated several times until the hammer comes to rest. The same thing occurs in a construction like that of fig. 4; before the contact between C and D is "definitely" made it is broken several times by the rebound of D. The result is that the current through the magnet coil is interrupted a great many times more than is necessary, which has an unfavourable effect on the life of the contact.

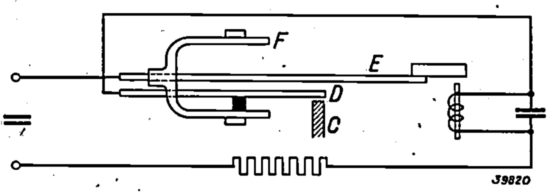


Fig. 5. The contact is made between D and F. This contact is closed at a moment when D does not rest upon C.

The principle of the improved construction is sketched in *fig. 5*. C is here again the stop for the spring D. Now, however, the spring D is not driven by the electromagnet but by the spring E to which the fork-shaped part F is fastened. The contact K is formed between this fork-shaped part and the spring D. The contact is closed when the spring E and thus also F moves upward, *i.e.* when D does not touch C. This construction may be compared with a hammer which strikes against an easily movable object; no rebound of the hammer takes place. The dying out of the vibration of the contact is thus prevented by this construction and a much longer life of the contact is obtained.

Improved construction of the contacts K_{11}-K_{22}

In the case of the contacts which serve for the continual alternation of the direction of the current also (fig. 1: K_{11}, K_{12}, K_{21}, K_{22}) the above-mentioned undesired damped vibration often occurs. The principle of the method by which this phenomenon has been combatted in these contacts is most clearly demonstrated by a comparison with billiard balls. If the ball *1* in *fig. 6*, moving in the direction of the arrow, strikes the stationary balls *2* and *3*, *1* and *2* remain motionless or move only very slowly further, while *3* rolls on alone. The same thing happens with the springs of the vibrator. As indicated in *fig. 7* behind the so-called side springs *2* are the extra springs *3*. When *2* is struck by *1*, *1* and *2* remain motionless, or move together over a short distance, and the contact between *1* and *2* is not broken. The spring *3* moves away from *2* but this causes no breaking of the contact.

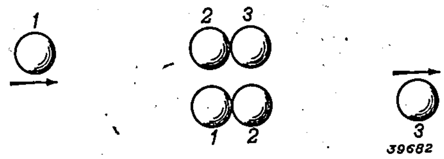


Fig. 6. When ball *2* is truck by ball *1*, *1* and *2* remain stationary, while *3* rolls away alone.

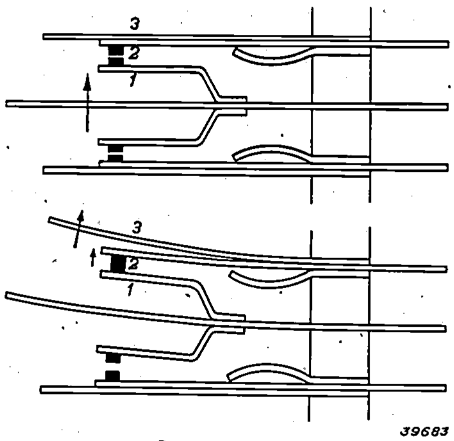


Fig. 7. By the action of the spring *3* the contact between *1* and *2*, when once made, is prevented from being broken several times by the damped vibration. The distance between the contacts is exaggerated in this figure.

Limitation of the current

During ordinary use the vibrator takes up a current from the D.C. mains which is given by the load on the A.C. end. Under certain circumstances, however, this current may temporarily assume a much too high value. This may take place for example when a small particle of metal has fallen between one of the contacts K_{11}-K_{22}. Since the distance between the contact surfaces in the open state is only 0.2 mm, a very small conducting particle is enough to cause one of the contacts to remain closed. The primary winding of the mains transformer of the radio set is then connected to the D.C. mains for too long a time and the current takes on too high a value. In order to prevent injury to the vibrator or the mains transformer, a fuse is included in the connection between the two. Since a defect of the nature described is generally of a temporary nature, the melting of this fuse is often not desired. For this reason a maximum relay is included in the vibrator which interrupts the connection with the contacts K_{11}-K_{22} upon the occurrence of too large a current impulse. The vibrator continues to function, but an unnecessary melting of the fuse has been prevented. Upon switching on the apparatus also a large current surge may occur, whereupon the maximum relay also goes into action.

Three coils are connected to the maximum relay, indicated by S_1, S_2 and S_3 in *fig. 8* in which the diagram of the complete vibrator is drawn (for

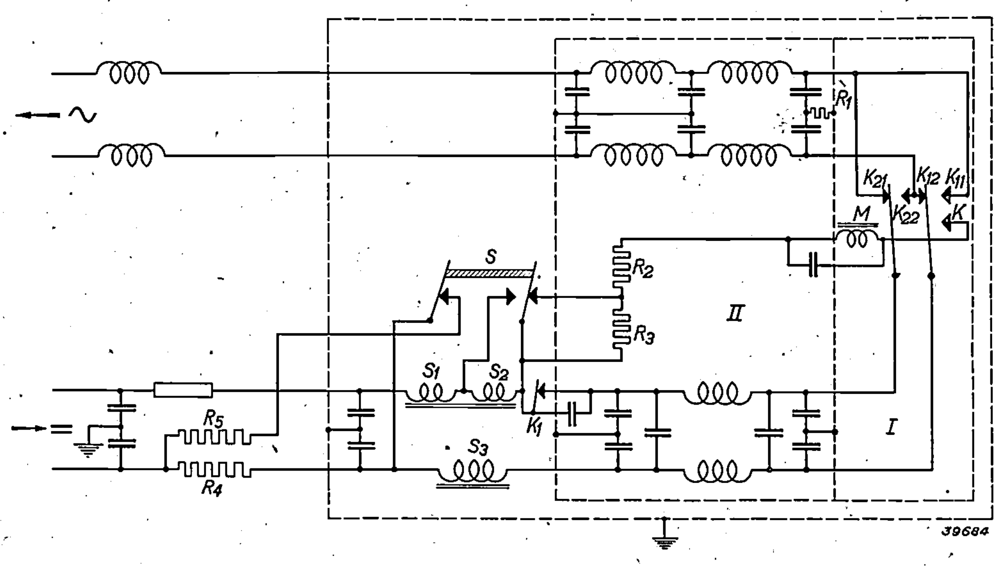


Fig. 8. Complete diagram of the connections of the vibrator. The vibrator proper and the anti-interference part are each placed in a separate shielded compartment, indicated by *I* and *II*, respectively. Outside of these are the maximum relay and an arrangement for switching over the vibrator for D.C. mains with different voltages. Around this whole the second shielding, already referred to, is placed. The vibrator can be connected to D.C. mains with a voltage of 220 volts as well as to those of 110 volts. For switching from 220 V to 110 V, three components must be changed:

1) A resistance in series with the magnet coil (R_2, R_3). At the low mains voltage the resistance R_3 is short-circuited.
2) The maximum relay. Since the vibrator takes up a larger current at a lower mains voltage, the maximum relay must come into action at a higher current value. This is achieved by connecting the coil S_2 at the low mains voltage, which coil is wound in an opposite direction to S_1.
3) The resistance which is connected in series with the vibrator. Upon use of a low mains voltage the resistance R_4 is connected in parallel with R_5.

The commutation arrangement is indicated by the double switch *S*, Actually this is a contact block on the outside of the vibrator, which only needs to be reversed for switching over to the other mains voltage.

the significance of S_1 and S_2 see the text under the figure). When the current exceeds a certain maximum the contact K_1 is broken. This contact is also shunted by a condenser as may be seen in the diagram. In order to limit possible current surges a small resistance R_4 is connected in series with the whole. This resistance is situated in the connection of the vibrator leading to the radio set.

Complete shielding

It is always very dificult to house an apparatus in which the heat development is fairly high in an entirely closed shielding container. This is, however, very desirable for the vibrator, in connection with the great interference which it may otherwise cause in the radio reception. We have succeeded in keeping the energy loss which occurs in the vibrator so low that the apparatus could be placed in a completely closed container, thus even without ventilation holes. In order to obtain the greatest possible security from interference the shielding is double. The energy loss was kept small, and thus the heat development also, by making the moving parts (springs and armature) as small as possible. By this means a small current, namely only 20 mA, in the coil of the electromagnet M is sufficient to keep the vibrator in motion.

Keeping the energy loss small is of course also an advantage from the point of view of efficiency. In the case of this vibrator the efficiency amounts to about 90 per cent.

Finally in *fig. 9* the complete vibrator is shown, The double shielding of the vibrator proper has here been removed in order to show the main parts more clearly.

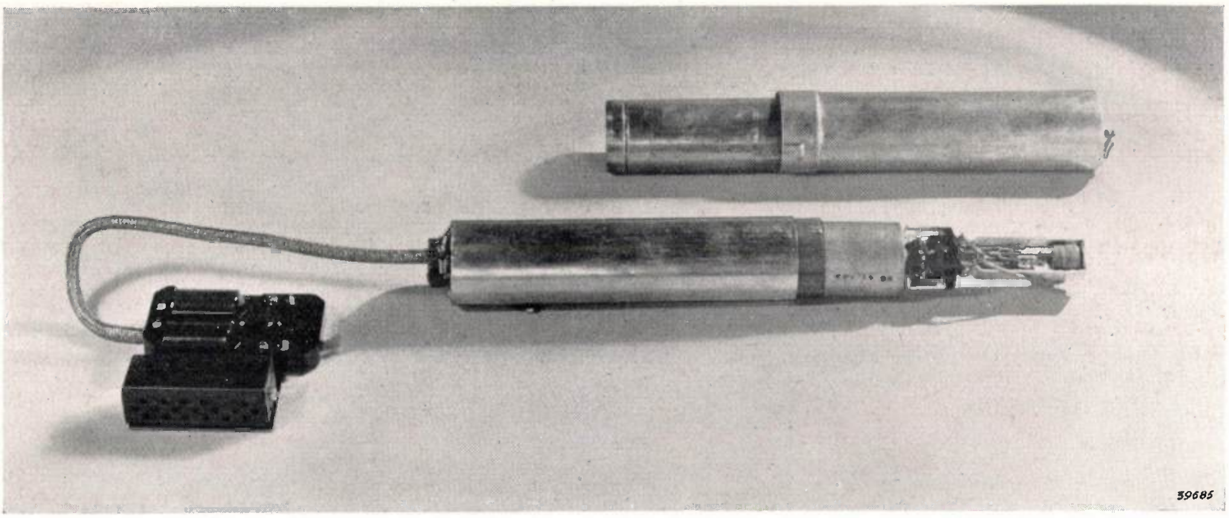


Fig. 9. The complete vibrator. Left, anti-interference part, right vibrator. Behind, slid partially into one another, are two cans with which the vibrator is doubly shielded.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

1545: J. L. Snoek: The influence of eddy currents on the apparent hysteresis loop of ferromagnetic bars (Physica 8, 426-438, Apr. 1941).

In this article a commonly neglected error in measurement is discussed which occurs in ferromagnetic measurements on bars or wires of magnetically soft material. In this case the external magnetic field is practically completely compensated by the demagnetizing field. If the external field is now changed so rapidly that the field in the interior of the bar can only follow with a large time time lag due to eddy currents, the demagnetizing factor is temporarily very much reduced. This results in the fact that the interior of the rod is much too strongly magnetized for a short time. When the final state is once reached the inner part of the bar is thus on the descending branch of the magnetization curve instead of on the mounting branch. The area of the hysteresis curve thus appears smaller than it actually is. The coercive force measured is also too small. These errors in measurement occur when the time constant of the external field is small compared with the time constant of the material. The magnitude of the error in the change in induction observed also depends upon the ratio of the field change to the accompanying change in the demagnetization field. The error therefore does not occur in ring-shaped test pieces which have no poles. The shape and the magnetic properties of the material further determine the maximum error which can occur. With materials with not too small a coercive force the occurrence of this error need not be feared very much, since in this case the time constant of the external field will usually be much larger than that of the material.

1546: M. J. Druyvesteyn: The elastic anisotropy of molybdenum (Physica 8, 439-448, Apr. 1941). (Original in German).

The modulus of elasticity and the modulus of torsion are determined for polycrystalline molybdenum and for strips which have been cut with a chisel in different directions out of rolled plates of molybdenum, which clearly show a texture. The different elastic constants of molybdenum can now be calculated from these measurements and the compressibility determined by Bridgman. Molybdenum is the first metal for which it has been determined that the modulus of elasticity in the direction of the main diagonal is smaller than that along the edges of the elementary cube of the crystal.

1547*: W. G. Burgers: Recrystallization, deformed state and recovery (553 pp; Akad. Verlagsgesellschaft, Becker und Erler Kom. Ges., Leipzig 1941) (In German).

This book is published as volume III, part II of the Handbuch der Metallphysik and offered the writer the opportunity of presenting his extensive knowledge of the subject of recrystallization in a comprehensive way. After an introduction in which, among other features, a nomenclature for the different phenomena of recrystallization is given and the methods of investigating them are dealt with, follow chapters on recrystallization in unworked substances, on the cold-worked state and its recrystallization. Furthermore the recrystallization temperature and the duration of recrystallization are discussed, as well as the influence of impurities and alloy components on the recrystallization of purer metals. In conclusion the significance is discussed of recrystallization on stiffening and plasticity.

1548: M. J. O. Strutt: Spontane spannings- en stroomfluctuaties (ruischen) in electronenbuizen en aangeloten ketens (Spontaneous voltage and current fluctuations (noise) in electronic values and circuits connected with them) (T. Ned. Rad. Genoot. 9, 1-36, June 1941).

In this lecture given before the Netherlands Radio Society (Nov. 1940) a survey was given of the noise phenomena which have in general already been dealt with in detail in the different articles devoted to "noise" in Philips techn. Rev. Equivalent circuit diagrams and a short derivation are given which constitute a considerable simplification in the treatment of noise phenomena.

1549: B. D. H. Tellegen: Meetkundige configuraties en dualiteit van electrische netwerken (Geometrical configurations and duality of electrical networks) (T. Ned. Rad. Genootschap **8**, 37-60, June 1941).

This lecture given before the Netherlands Radio Society (Nov. 1940) is an elaboration of the article contributed by the author to Philips techn. Rev., **5**, 324, 1940. He took the opportunity of going somewhat more deeply into the proofs of the theorems dealt with and of illustrating them by means of a few examples.

1550: F. A. Heyn: Het opwekken van mutaties door straling (The excitation of mutations by radiation). (Vakbl. Biol. **22**, 81-88 and 101-105, May and June 1941).

The mutation-causing effect of different kinds of radiation is investigated as a function of the dosage and other external conditions. It is found that there is no question of a specific action of certain kinds of radiation, but that mutations occur when sufficient energy is transferred to a certain sensitive part of a gene. It is not possible to excite definite desired mutations in this way, but there is a good chance of encountering useful mutations among the many less vigorous ones which are obtained with a sufficiently intense radiation, so that in addition to theoretical significance for the general study of heredity, this method will also certainly have practical significance for the improvement of plants and animals.

1551: J. van Niekerk: On the influence of emanation on the development of experimental rickets in young rats (Act. brevia Neerl. **11**, 142-147, 1941).

The large fluctuations which occur in the development of artifically produced rickets in large numbers of rats, which are needed for instance for the standardization of vitamin D preparations, are ascribed by various investigators to the variability of different external factors. Even though great care is taken to secure constancy of the rachitogenic diet fed to the rats, the large fluctuations in the development of the rickets continue to occur. Dols and Jansen ascribed this to variations in the content of radium emanation in the air with the season of the year. By keeping different rats under otherwise similar conditions in cages with different contents of radium emanation the author has shown that this content has no effect on the development of experimental rickets in rats.

1552: G. W. Rathenau and J. L. Snoek: Magnetic anisotropy phenomena in cold-rolled nickel-iron (Physica **8**, 555-575, June 1941).

The strong magnetic anisotropy observed in cold-rolled nickel-iron alloys can neither be explained as due to internal stresses nor as due to the natural anisotropy of the crystals. A geometric tetragonality of these alloys which was indicated by G. Wassermann as the cause, cannot fully explain the phenomenon. The influence of the composition of the alloy and of the treatment which it has undergone on the anisotropy of its mechanical properties was further investigated. In this investigation it is particularly remarkable that the natural crystal anisotropy and the magnetostriction are not appreciably affected by the rolling. The investigations suggest a relation between the anisotropy phenomena and the familiar phenomenon of order in the nickel-iron system. The authors, however, do not consider it very probable that the anisotropy phenomena are a direct result of the tendency toward order.

1553: M. J. O. Strutt and A. van der Ziel: Decrease and elimination of the spontaneous fluctuations in the amplification of extremely small photocurrents (Physica **8**, 576-590, June 1941) (Original in German).

In continuation of **1529** the authors discuss the amplification of extremely small photocurrents in connection with the fluctuations occurring in the amplifier. These fluctuations can be limited by using only a narrow frequency band preceding the first amplifier stage, and extending it in the following stages, or by making use of an electron multiplier. A new method of limiting the fluctuations is indicated in which a special method of backcoupling is applied. The ratio of signal to noise is calculated and compared with that in the previous cases. In conclusion the extent is examined to which this new method for limiting the noise can also be profitably employed in other cases.

Contents of Philips Transmitting News **8**, No. 2, June 1941.

VOL. 6 No. 12, p. 349-376 DECEMBER 1941

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
N.V. PHILIPS' GLOEILAMPENFABRIEKEN

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

LUMINESCENT SUBSTANCES

by F. A. KRÖGER. 335.37

On the basis of the atomic theory a concept is developed of the mechanism of the fluorescence and phosphorescence of the luminescent substances commonly used in technology, such as zinc sulphide, willemite, calcium and magnesium tungstate, cadmium borate, etc. Especial attention is paid to the energy levels of the electrons in a solid luminescent substance and their relation to emission and absorption phenomena.

Introduction

At the present time a large number of luminescent substances are known. These are substances with whose help it is possible to convert corpuscular radiation, X-radiation or ultra violet radiation into light, or in general into a form of radiation which is different from the original form.

To begin with a simple example of the application of such substances we may refer to the luminous dials of watches which are covered with a preparation which contains in addition to the luminescent substance a very small amount of a radioactive element (radium or mesothorium). The latter elements emit alpha particles of high energy which cause the luminescence of the preparation.

A second example is formed by the screens which are used in fluoroscopy with X-rays. The screens have the property of emitting intense light under the influence of X-radiation, so that the shadow picture thrown by the X-rays on the screen becomes visible to the eye. A related application is formed by the reinforcing screens for X-rays which produce photographically active light, so that a more intense image is formed than upon direct action of the X-rays alone.

The conversion of electron rays into light is encountered in the case of the substances from which the screens of cathode ray oscillographs and television tubes are made. Finally, the conversion of ultra violet radiation into light takes place in modern gas-discharge lamps (luminescence lamps).

Fluorescence and phosphorescence

The conversion of electromagnetic (especially ultra violet) radiation into light of another wave length, namely into visible light, is called fluorescence [1]). In addition to the phenomenon of fluorescence, many fluorescent substances exhibit a second peculiarity, namely a very long after-luminescence (very familiar at present in the case of the luminescent buttons and pins which are worn during the blackout). This phenomenon, phosphorescence [2]), is distinguished from fluorescence by its great dependence on temperature. If a luminescent button, after having been irradiated, is immersed, in the dark, in a glass of hot water, it suddenly gives much light, which, however, rapidly decreases in intensity. Conversely, a low temperature has a retarding action on the giving off of the energy, so that the luminescence lasts longer when it is cold. At a very low temperature (immersion in liquid air) the luminescence disappears entirely; the energy absorbed can then be stored for an almost unlimited time and whenever desired it can be produced again by heating. The mechanism of phosphorescence is closely connected with that of fluorescence, but it must be considered separately. A single word which is used to indicate both the phenomena is luminescence. For that reason the fluorescent and phosphorescent substances used in technology, especially when the phosphorescence is only a subsidiary property, are often called luminophors.

On previous occasions the phenomenon of

[1]) From the mineral fluor-spar or fluorite (CaF_2) which in the presence of certain admixtures, such as some of the rare earths, clearly shows the phenomenon.

[2]) Phosphor = light bearer, is the name of these preparations originating about the year 1600, which then first became known in Europe.

fluorescence has been studied in some detail in this periodical, gases and liquids as well as solid substances were then discussed. In this article we shall go somewhat more deeply into the luminescence of solid substances. We shall discuss, in addition to the chemical composition, the mechanism of light emission in the case of free atoms and crystals, the absorption spectrum of the substances in question, the connection between absorption of light and fluorescence, and, further in particular the influence of small amounts of foreign substances on absorption and emission.

Chemical composition

Luminescent substances of very divergent chemical composition are known. Certain substances (salts of rare earths, uranyl salts, benzene) possess the property of luminescence in the pure, unmixed state. In other cases the phenomenon only becomes intense when the substance in question is present as a slight admixture in another substance. Ruby is an example of this. Lecocq de Boisbaudran (1886) showed that Cr_2O_3 as an admixture in the basic material Al_2O_3 is the active component. A second example of an admixture phosphor, which also exhibits a strong phosphorescence, is CaS. This compound forms the chief component of the luminescent paint whose luminescence was shown by Verneuil (also in 1886) to be due to extremely small traces of bismuth (Bi_2S_3). In the same year E. Becquerel discovered that calcite ($CaCO_3$) only exhibits luminescence in the presence of small amounts of manganese (Mn). Becquerel and Lenard (1890) carefully examined the properties of such preparations.

In addition to the above-mentioned compounds of rare earths a number of other substances are known which are luminescent even when no foreign substances are present. These are substances in which a deviation from the stoichiometric chemical composition is the condition for the appearance of the phenomenon. Thus ZnS and ZnO are luminescent when a small excess of Zn-atoms is present in the lattice. Somewhat similar cases are those of $ZnCO_3$ and $CaHPO_4$, where it is known that they only exhibit luminescence in a partially decomposed state.

The mechanism of the emission of light

According to the quantum theory a system of electrons, such as occurs in the case of an atom, molecule or crystal lattice, can only exist in discrete states, each of which is distinguished by a definite energy content. We thus distinguish the following energy states: E_0, E_1, E_2, etc. The state with the lowest energy (E_0) is called the basic or normal state. When the system is in a state of higher energy (also called an excited state) in many cases it returns spontaneously to a lower state, emitting as it does so a quantity of light (light quantum) such that Bohr's (1913) equation is satisfied:

$$E_2 - E_1 = h\nu_e.$$

In this equation ν_e represents the frequency of the light emitted (in sec.$^{-1}$); while h is a constant ($h = 6.6 \times 10^{-27}$ erg sec.). If the energy is measured, as is often done, in electron volts instead of ergs ($1\ \text{eV} = 4.8 \times 10^{-10} \times 1/300 = 1.6 \times 10^{-12}$ ergs) and if, instead of the frequency, the number of wave lengths per cm ($\nu_{\text{cm}^{-1}}$) or the wave length λ in Å is used, then

$$(E_2 - E_1)_{\text{eV}} = 8072\ \nu_{\text{cm}^{-1}}$$
$$(E_2 - E_1)_{\text{eV}}\lambda_{\text{Å}} = 12\,390.$$

In general it is possible to bring a system into a higher state of energy by irradiating it with corpuscular rays or with light. In the latter case the following condition must be exactly satisfied:

$$E_a - E_0 = h\nu_a,$$

where E_0 represents the energy of the basic state and E_a that of the excited state. ν_a is the frequency of the light which causes the excitation.

If the system is brought into the excited state by corpuscular rays (charged particles) the following less stringent condition is then in general valid:

$$\tfrac{1}{2}\, mv^2 \geqslant E_a - E_0,$$

where $\tfrac{1}{2}\, mv^2$ is the kinetic energy of the particle of mass m and velocity v.

If the system being considered is a free atom or molecule, E_a simply represents one of the states in which one of the electrons is displaced to an orbit with a higher energy. In the case of a crystal which is excited by light the same is true.

If excitation is effected by corpuscular rays (cathode rays, α-particles, electrons freed by X-rays) it must, however, be supposed that with the retardation in the crystal a single incident particle causes a large number of electrons to shift to higher orbits.

The process of emission is the same in both cases. For the sake of simplicity we shall in the following discuss only the luminescence caused by the absorption of light.

Some remarks on atomic structure

According to modern conceptions an atom consists of a positive nucleus surrounded by a more or less spherical swarm or cloud of electrons. According to the original concept of Bohr, neglecting the changes which the quantum theory has undergone since 1925 (wave mechanics), each electron describes a definite orbit. As an example in *fig. 1*

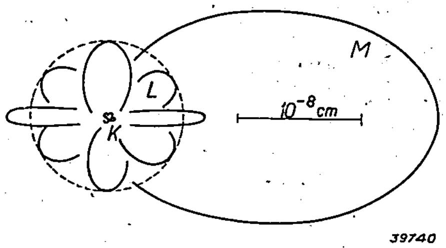


Fig. 1. Diagram of a sodium atom with its electron orbits according to Bohr's theory (from the year 1915).

a representation of a sodium atom according to Bohr's theory is given [3]). Each type of orbit can be traced by only a limited number of electrons (Pauli principle). Some orbits remain very close to the nucleus. These thus contain strongly bound electrons. Other electrons move farther away from the nucleus and are thus less strongly bound. The orbits can be classified into shells. The K, L, M, etc. shells can be distinguished, with the strength of bond decreasing in this order, so that the K-electrons are more strongly bound than the L-electrons, the latter more strongly than the M-electrons, etc. The maximum occupation of the shells is:

K	L	M	N
2	8	18	32

According to the nature of the distribution of the electrons over the orbits the atom may exist in different states of energy, as indicated above. It is now found to be permissible to a certain extent to state that each electron moves in the average electric field of the nucleus and of all the other electrons together. One may therefore speak of the energy (sum of potential and kinetic energy) of one electron in its orbit, and the question may be put as to the magnitude of this energy. An estimation of this can be obtained from the so-called ionization voltages, which are the amounts of energy, expressed in eV, which are necessary to remove one electron from the atom. These energies can be determined by the methods of ordinary spectroscopy and X-ray spectroscopy or with the electron impact method (Franck and Hertz 1914). In the case of sodium, for example, the following results are obtained.

The sodium atom contains a nucleus with the charge $+ 11e$ (e = charge on an electron $= 4.8 \times 10^{-10}$ e.s.u.) and 11 electrons. Two of these are bound in the K-shell with an energy of about 1000 eV, 8 in the L-shell with about 30 eV and 1 in the M-shell with about 5 eV. The further possible M-orbits and all N and higher orbits are empty in the normal sodium atom. It is found to be possible under the influence of light absorption or collision with a foreign particle that one of the electrons may leave its normal orbit and go into one of the unoccupied orbits or be removed from the atom. Thus for example the sodium atom may be in a state in which the M-electron travels in an orbit which, from the point of view of energy, is about 2 volts higher than the orbit in which it normally moves. In *fig. 2* the energies of the different electron orbits of sodium are given schematically, and in *fig. 3* the energy states of the atom considered

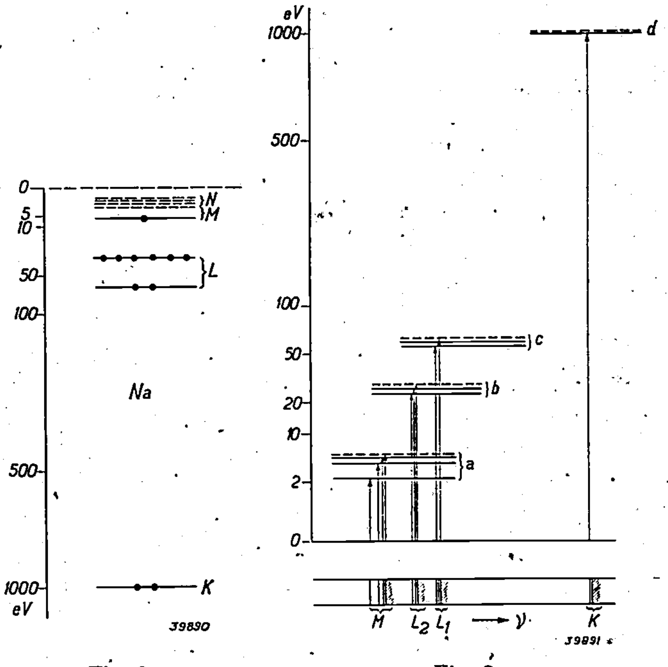


Fig. 2 Fig. 3

Fig. 2. Energies of the occupied (full lines) and unoccupied (broken lines) electron orbits in the sodium atom. In each case as zero point the state is chosen in which the electron is at rest at an infinite distance from the atom.

Fig. 3. Energy states of the sodium atom which result from the displacement of one of the electrons to orbits with higher energy. (The ordinate is drawn proportional to the square root of the energy).

States *a* result from a displacement of the M-electron, *b* and *c* result from excitation of one of the L-electrons, respectively, and *d* from excitation of a K-electron. As zero level the energy is chosen of the basic state of the atom given in fig. 2. Below may be seen a schematic representation of the absorption spectrum (on a non-linear scale).

[3]) Taken with slight alterations from: H. A. Kramers and Helge Holst, Bohrs atomtheori, Copenhagen 1915. The figure is slightly modified in accordance with the improvements which were introduced in the theory by Pauli, Stoner and Hund.

as a whole, which occur when one of the electrons passes from its original orbit to one of the unoccupied orbits whose energies are indicated in fig. 2 by dotted lines. Such transitions caused by absorption of radiation (light or X-rays) are indicated by arrows. The absorption spectrum is shown below the diagram of energy levels. This spectrum consists of a number of groups of lines which converge toward a limiting frequency which forms the beginning of a region of continuous absorption (absorption edge).

In addition to relatively simple cases like that of sodium there are also more complicated ones. In fig. 2 it was indicated without further remark that the L-shell contains two groups of orbits with different energies. The complete M-shell contains 3 subgroups, the complete N-shell 4 subgroups, etc. The maximum occupation of these subgroups by electrons is the same for each shell, beginning with the lowest it is always 2, 6, 10, 14 . . . . . Let us now take as an example of a complicated case the triply ionized atom of chromium (Cr^{3+}) which causes the luminescence of ruby. In the case of this ion the K and L-shells are completely occupied by electrons, likewise the first two subgroups of the M-shell. The third subgroup, however, has only 3 of the 10 electrons which it should contain at the

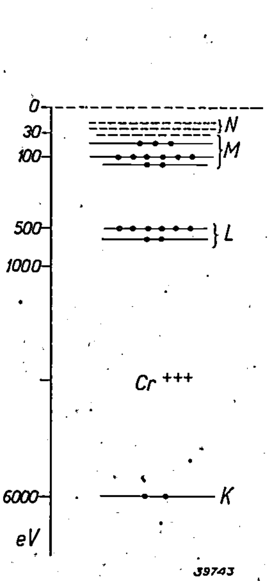


Fig. 4

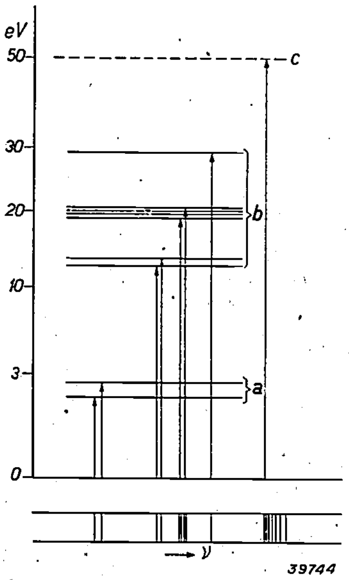


Fig. 5

Fig. 4. Energies of the electron orbits in the triply ionized chromium atom.

Fig. 5. Energy states of Cr^{3+} (non-linear scale).
a states occurring due to the regrouping of M-electrons,
b states occurring due to displacement of an M-electron to an N-orbit.
c ionization limit of the M-electron.
States which occur due to the displacement of an L or K-electron are not indicated. Below the energy scheme is the corresponding absorption spectrum (schematic).

maximum. In *fig. 4* this state is shown as nearly as possible to scale.

The chromium ion can also pass into an excited state in which the energy content is greater than normal. Disregarding X-ray absorption, this may happen in the first place due to the fact that one of the most weakly bound electrons of the M-shell moves to an unoccupied higher orbit (for instance an N-orbit). There are, however, also states with an energy content greater than normal which occur due to the fact that the 3 electrons in the most weakly bound M-orbits are relatively differently grouped as regards their manner of movement. In *fig. 5* a number of these energy states of the chromium ion are represented: at *a* the states which occur by regrouping in the M-shell and at *b* the states which occur when an electron from the M-shell is brought into an N-orbit. Since such energy states often lead to absorption in the visible region of the spectrum, the ions with incompletely occupied subgroups are also called c o l o u r e d ions.

Energy states of crystals

The state which exists in a crystal whose components (atoms or ions) are regularly arranged at distances of about 10^{-8} cm apart can best be approached by beginning with a sort of expanded crystal lattice in which the components are already placed according to the symmetry of the crystal in question, but in which the distances are so great that the ions or atoms do not yet affect each other. Let us examine for instance the energy scheme for a crystal which is built up of univalent positive ions A^+ and univalent negative ions B^-. For the energies of the individual electron orbits the considerations of the preceding section then hold. In addition to transitions of an electron to a higher orbit within the same ion, transitions can now also be imagined in which an electron passes from one ion to another. The charges of the ions involved are thereby changed. For example a positive ion A^+ which receives an electron becomes a neutral atom A, a negative ion B^- upon taking up an electron becomes a doubly charged ion B^{2-}. If we wish to represent these transitions in a scheme of electron energies, then in addition to the electron states of the ions A^+ and B^- we must also include those of A and B^{2-}. This has been done in *fig. 6a* in which, however, representation to scale has not been considered and no particular ions were in mind.

The position of each level is determined by the energy which is necessary to bring an electron from the level in question of an atom or ion to infinity. In connection with this definition, in the indication of a transition between two levels the nomenclature of the level which g i v e s up the electron offers no difficulty. If, however, the electron is taken up by an ion (A+ for instance), we must give to the unoccupied level in question of this ion the name of the neutral atom formed (A). By definition the position of this is given by the energy which is necessary to take an electron from this neutral atom to infinity. This is, however, just the opposite of the energy which is received when we make an atom out of the ion by adding an electron. The transition of an electron from an ion B− to an ion A+ is thus given in the energy scheme by the transition from the level B− to the level A.

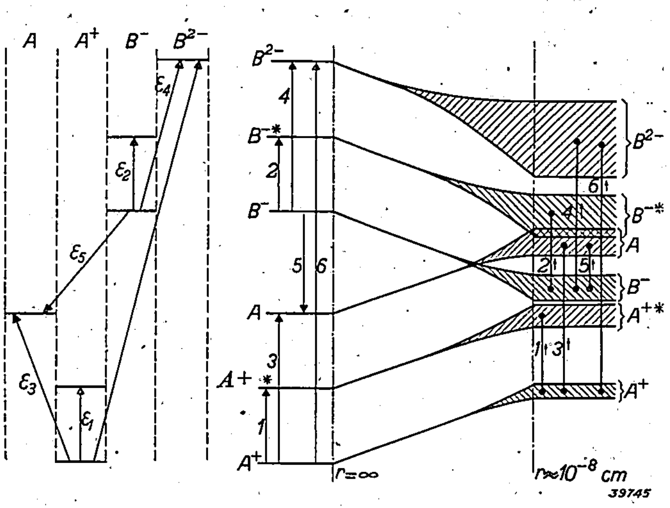


Fig. 6. Scheme of the possible energy states of the most weakly bound electrons of a positive ion A+, a neutral atom A and negative ions B− and B2−: *a*) in the free state, *b*) combined to form a crystal.

Possible transitions between the levels of fig. 6 are:

$$A^+ + \varepsilon_1 \rightarrow A^{+*}\ ^{4)}$$
$$B^- + \varepsilon_2 \rightarrow B^{-*}$$
$$A^+ + A^+ + \varepsilon_3 \rightarrow A^{2+} + A$$
$$B^- + B^- + \varepsilon_4 \rightarrow B + B^{2-}$$
$$A^+ + B^- \rightarrow A + B + \varepsilon_5$$
$$A^+ + B^- + \varepsilon_6 \rightarrow A^{2+} + B^{2-}$$

It will be seen that in the fifth process energy is liberated. This shows clearly that positive and negative ions are not stable side by side if there is no interaction between them. This may be understood when it is kept in mind that the formation of a positive ion from a neutral atom usually requires more energy, than is liberated when an electron is added to a neutral atom forming a negative ion (there are many cases indeed in which this latter process also requires energy!).

Let us now allow the ions of the expanded lattice to approach each other until a crystal is formed in which the ions are at normal distances apart (about 10^{-8} cm). Between the ions electrostatic forces of attraction then begin to act. In the final state, the normal crystal, every positive ion is surrounded by a number of negative ions (in NaCl the number is 6), and every negative ion by a number of positive ions. At a greater distance there are indeed ions of the same sign as the ion under consideration, but the result will nevertheless be that the potential energy of the electrons at the position of a positive ion has risen and at the position of a negative ion it has fallen. As a result of this in our diagram the electron energy levels of A must rise compared with those of B−, and by an amount of the order of magnitude of 10 eV. This is shown schematically in fig. 6*b*. It will be seen that due to the mutual attraction of the ions the level B− is now below A [5]).

In addition to the change in relative position of the levels, a second effect is also indicated in fig. 6*b*, namely a broadening of the levels. This effect is of a more complicated nature and was only explained by the modern atomic theory (wave mechanics). It is connected with the fact that upon approach of the ions the electron orbits partially overlap. Each level thereby splits up into a very large number of sublevels, namely as many as there are ions of the same kind in the crystal. Each sublevel can in general only be occupied by two electrons at the most. The collection of all sublevels belonging together which have been formed from one atom level is called an energy band. From what has been said about the cause of the broadening it also follows that it will be greatest for those electron levels for which the corresponding orbits extend farthest outside the ion. The levels of the most strongly bound electrons, which remain very close to the nucleus, are thus only slightly broadened. The same is true for the more weakly bound electrons if the shape of the orbit is of such a nature that the electron does not move especially far outside of the circumference of the ion. This is the case for example for the most weakly bound M-electrons of the Cr^{3+} ion discussed above, and for the most weakly bound N-electrons of the rare earth atoms. If these atoms are of the coloured type, in which different energy states are possible due to different orientations of the orbits of the M or N-electrons, with respect to each other or to the orbits of the remaining electrons, the situation is one in which the

[4]) An asterisk here indicates an excited atom or ion.

[5]) The figure is only a sketch and therefore more or less arbitrary. There are indeed cases in which the "crossing" of A and B− would not take place. This means that the starting point was wrong and that one was not dealing with an ionic crystal but with a crystal lattice which consists of atoms which attract each other by so-called homopolar binding forces.

transitions between these energy states take place as it were inside the atom (ion) without the whole crystal as such taking part. In *fig. 7a* the system of electron levels is given schematically for the combination: coloured positive ions and non-coloured negative ions, and in fig. 7*b* the energy band system of a crystal built up of these ions. Negative coloured ions also exist (NO_3^-). In general in this case the isolation of the electron orbits in question is less pronounced, so that the energy levels are broadened into bands.

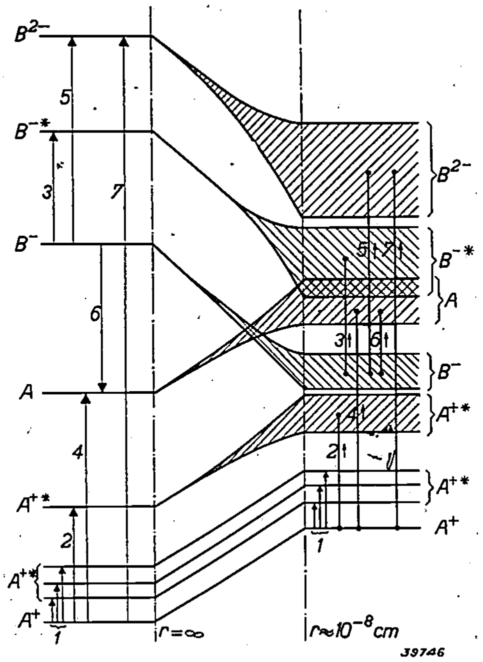


Fig. 7. Like fig. 6, but for the case in which the ions A+ are "coloured" ions.

Absorption spectrum of crystals

When a crystal is irradiated with light, then, analogous to what we have seen in the case of atoms and ions, absorption will only take place when the size of the quantum $h\nu$ corresponds to the energy difference between an occupied and an unoccupied band.

Since the position of the bands depends upon the nature of the ions, the crystal structure and the nature of the crystal lattice, these factors will also determine which transitions require the least energy.

In a normal ionic crystal the absorption spectrum will usually have the appearance of that sketched in *fig. 8* (corresponding to the case of fig. 6). The lowest absorption frequency here corresponds to the transition indicated by *5* of an electron from the band of the negative ions to that of the positive ions. All the other absorption bands lie at higher frequencies, more or less overlapping.

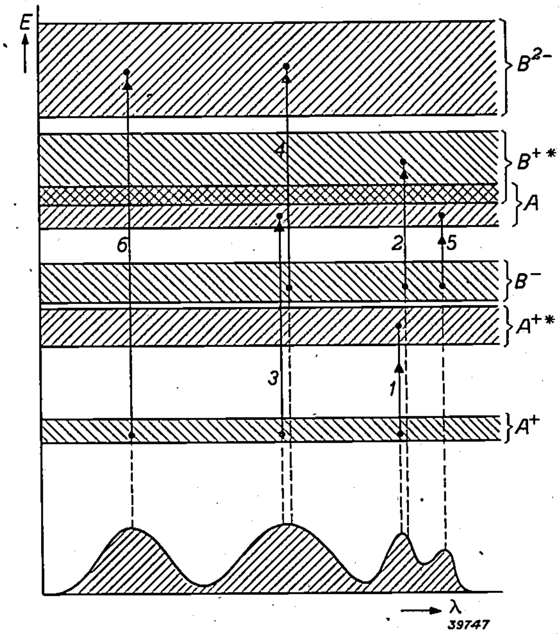


Fig. 8. Energy bands of the electron states of the crystal of fig. 6, with the absorption spectrum below.

For a crystal among whose components are coloured ions, the absorption spectrum is given in *fig. 9* (corresponding to fig. 7). In addition to broad absorption bands for which it is assumed that the band with the lowest frequency again corresponds to the transition from B^- to A^+ (with the formation of A), we now also encounter more or less sharp absorption lines, which are the result of internal transitions in the coloured ions. In the

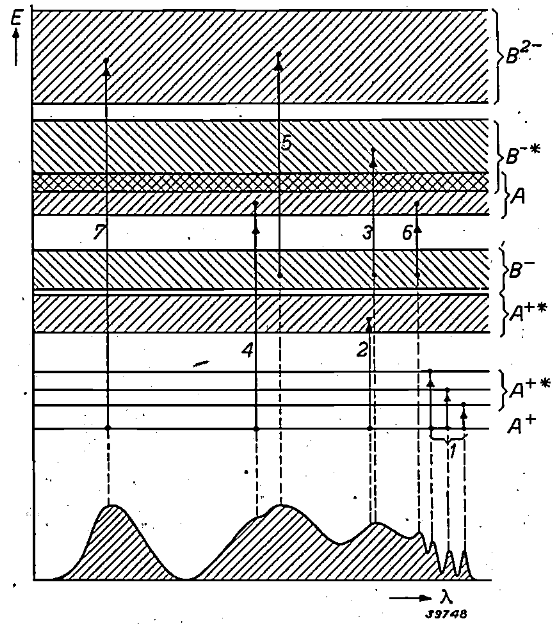


Fig. 9. Energy bands and absorption spectrum of a crystal with "coloured" positive ions.

scheme of fig. 9 it is assumed that these absorption lines lie at the long-wave edge of the crystal absorption bands (this is the case with most salts of the rare earths, *e.g.* $SmCl_3$). It will be clear that it need not always be so. In those cases, however, where the sharp absorption lines of the ions fall in the region of crystal absorption they cannot of course be observed. In the case of coloured negative ions also separate absorption bands of these ions are often observed at the long-wave end of the spectrum. Thus $CaWO_4$ and $NaNO_3$ exhibit absorption bands in the near ultra violet, which are to be ascribed to electron transitions within the WO_4^{2-} and NO_3^- groups, respectively, while the real crystal absorption bands of these crystals lie in the short-wave ultra violet.

Electron energies and absorption of disturbed crystals

As we saw at the beginning, in many cases luminescence occurs in connection with the defects in a crystal due to deviations from the stoichiometrical composition (unoccupied ion positions in the lattice and therefore local excess of the other kind of ion), or due to the fact that foreign ions are included in the lattice. The latter may occur in two different ways, namely:

a) due to the fact that foreign ions have taken the places of ions of the basic material (examples: ZnS in which Zn-ions are to a small extent replaced by Mn [6]) and Al_2O_3 in which several Cr-ions have occupied the positions of Al;
b) due to the fact that foreign ions have penetrated between the ions of the lattice (situated at so-called inter-lattice points; an example of this is ZnS to which a small quantity of Cu or Ag is added).

As a result of one of these two causes the energy bands in the immediate neighbourhood of these disturbed spots may undergo displacements as indicated in *fig. 10a*. As to the magnitude of these displacements, *i.e.* the local position of the energy levels with respect to the general band scheme, few prophecies can be made. We can, however, be certain in advance that we will obtain evidence of these extra levels only when they lie between the highest occupied band and the lowest unoccupied band of the undisturbed crystal, such as the occupied level D and the unoccupied level F in *fig. 10a*. In this case they may lead to transitions such as are indicated by *2* and *3* in *fig. 10a*. These transitions are observed in the absorption spectrum as an extension of the crystal absorption (transition *1* in *fig. 10a*) toward the long-wave end.

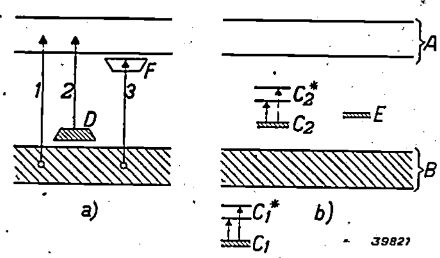


Fig. 10. *a*) Disturbance of the occupied and the unoccupied band, *1* crystal absorption, *2* transition from an occupied disturbance level to the empty lattice band, *3* transition from the full lattice band to an empty disturbance level.
b) Possible positions of the electron levels of built-in coloured and non-coloured ions. C_1, C_2 coloured ions, E non-coloured ions.

The foreign ions introduced or the excess ions present are also bearers of electrons which may again exist in different states to which (local) energy levels correspond. It is also impossible to say much in general about the position of these levels. It is, however, certain that the lowest occupied level must lie below the lowest unoccupied energy band of the crystal, since otherwise an electron would spontaneously pass from the foreign ion to the unoccupied band with the formation of a more highly ionized ion. If it is a question of foreign coloured ions, then in addition to the basic level and at a relatively small distance from it, other levels may occur corresponding to differently oriented orbits or to orbits of smaller bond energy, in the way discussed above for the case of Cr^{3+}.

In fig. 10*b* two possible positions of the energy levels for a foreign coloured ion are indicated by C_1 and C_2, and a possible position of a level of a non-coloured ion by E. In the qualitative consideration of the energy levels of foreign ions it makes little difference actually whether they form part of the regular lattice or have found a place in the inter-lattice points of the cristal or where the regularity of the crystal was already disturbed. Quantitatively there may be differences (height and relative position of levels) which we shall not go into here. If, as indicated in fig. 10*b* by C_1 and C_2, the distance between the levels of the foreign ion is smaller than the smallest distance between an empty and an occupied crystal band, more or less sharp absorption bands of the foreign ion can be observed at the long-wave end of the crystal absorption spectrum, just as in the case of a crystal which is built up entirely of coloured ions (fig. 9). In the case in which the foreign ions are

[6]) In contrast to the spectra of most other atoms and ions the spectrum of Mn^{2+} has only been investigated very incompletely. Therefore in fig. 3 Cr^{3+} was chosen instead of the much more commonly occurring Mn^{2+}.

situated at lattice points, and thus replace ions of the basic material, it may even occur that a gradual transition is possible from the case of fig. 10 to that of fig. 9, namely when a regular mixed crystal is formed in which the percentage of the foreign ions is gradually made to increase [7]). As an example of absorption by built-in coloured ions we give in *fig. 11* the absorption spectrum of the mixed crystal Zn_2SiO_4—Mn_2SiO_4 (willemite).

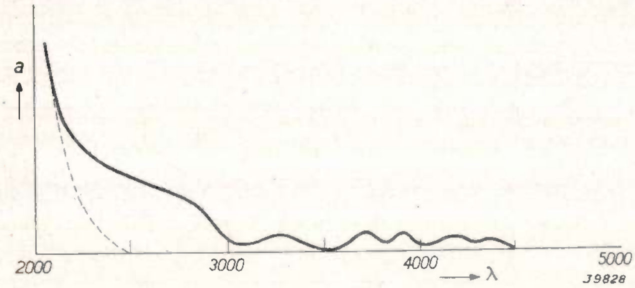


Fig. 11. Absorption spectrum of Zn_2SiO_4-Mn_2SiO_4 with 1% Mn.

The absorption is observed at $\lambda < 2500$ Å which is to be ascribed to the transition from SiO_4^{4-} to Zn^{2+}, at $2500 < \lambda < 3000$ Å the corresponding absorption band for the transition from SiO_4^{4-} to Mn^{2+} and finally at $\lambda > 3000$ Å various narrow and weak absorption bands which are to be ascribed to the Mn^{2+} ion itself. Since in all these wavelengthregions the absorption of light leads to fluorescence the absorption spectrum can also be made visible by means of fluorescence. *Fig. 12* shows such a picture.

Fig. 12. Excitation spectrum of willemite (Zn_2SiO_4-Mn_2SiO_4): *a* with 1 per mille Mn_2SiO_4, two different exposure times, *b* with 1% Mn_2SiO_4.
This spectrum was obtained by projecting the continuous spectrum of a hydrogen tube upon a thin layer of willemite which was separated from the sensitive layer of a photographic plate only by a green gelatine filter which only allows the fluorescence light of the willemite to pass.

The process of fluorescence

As was stated previously, in the case just considered of the crystals which contain foreign ions, fluorescence occurs as a rule, not only when they are irradiated with light of the wave length of the specific absorption bands of the ion (if such are present), but also when they are irradiated with light whose wave length lies in the region of crystal absorption. What idea must we form in this latter case of the mechanism of fluorescence? In the first place in the region of crystal absorption an electron is brought for instance from an ion B^- to an ion A^+ by the absorption of a light quantum, as is indicated in *fig. 13a* at *1*;

$$A^+ + B^- + h\nu \rightarrow A + B.$$

The electron then moves to one of the lower levels of the empty band (*2*), any excess of energy being transmitted to the crystal as thermal agitation, and can travel through the whole crystal. Another method of representation is that the electron in this energy state is passed on from ion to ion:

$$A_1 + A_2^+ + A_3^+ \rightarrow A_1^+ + A_2 + A_3^+ \rightarrow A_1^+ + A_2^+ + A_3 \rightarrow, \text{ etc.}$$

In an exactly similar way the empty place which has been caused by the neutralization of B^- will travel through the crystal, after (*3*) the empty level has moved upwards into the occupied band:

$$B_1 + B_2^- + B_3^- \rightarrow B_1^- + B_2 + B_3^- \rightarrow B_1^- + B_2^- + B_3 \rightarrow \text{ etc.}$$

In a completely undisturbed crystal two possibilities would now remain open for the electron:

a) return to the open place in the lower band with the radiation of light of about the same frequency as that of the absorbed light (*4*);
b) return as in a), but without radiation, the energy liberated being converted into vibrations of the crystal lattice (heat vibrations).

Both processes occur. Process *b*) is the normal conversion of absorbed light into heat. Process *a*)has been observed in certain crystals (CdS) at a low temperature.

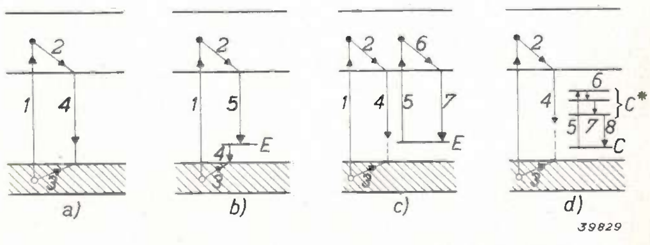


Fig. 13. Possible transitions in a crystal of the ZnS type.
a) without admixture;
b) with "non-coloured" ions built in;
c) idem, sensitized fluorescence;
d) like *c*, but for "coloured" ions.

If the crystal is disturbed by foreign ions (or by a stoichiometric excess of its own ions) and if energy levels lying between the two bands of the crystal (occupied and unoccupied) correspond to these ions, the emission of fluorescence light may take

[7]) A very good example is the absorption of mixed crystals of lanthanum fluoride (LaF_3) and cerium fluoride (CeF_3), whose spectrum contains characteristic absorption bands which are to be ascribed to the Ce^{3+}-ions and which upon increase of the cerium concentration gradually pass over into the characteristic absorption of pure CeF_3.

place in different ways. In the first place (fig. 13*b*), after an electron has passed from the full to the empty band (*1*, *2*, *3*) the empty place in the lower band may be filled by an electron from the occupied level E (*4*). The energy which is thereby liberated can return to the now empty level E with radiation (*5*).

In the second place it is possible that an electron from the empty band returns to the empty place in the occupied band, but in such a way that part of the energy is used to bring an electron from the level E into a higher level, while the excess is given off to the lattice as heat motion.

A similar process is known in the case of free atoms under the name of "sensitized fluorescence". In a mixture of mercury and thallium vapour, mercury atoms can be brought into a state 4.9 eV above the normal state by the absorption of light (λ = 2 537 Å). Upon collisions with foreign atoms these mercury atoms can pass into a state at 4.7 V in which the chance of returning to the normal state with light radiation is very slight ("metastable" state). If a metastable mercury atom now encounters a thallium atom, it transmits its energy to the latter, partly in the form of excitation energy and partly in the form of energy of translation (heat). The excited thallium atom then radiates in its own frequency, returning at the same time to the normal state.

In fig. 13*c* this process is represented for a single level (non-coloured ion), in 13*d* for a coloured ion.

No matter what process of emission is assumed to take place with the collaboration of foreign ions, it will always be essential for the process of emission that there be a collaboration of an electron (in the empty band) and an empty place (in the full band). It is for this reason that the process of emission in these cases shows similarity, as far as its progress with time is concerned, with a bimolecular chemical reaction. In an earlier communication this was dealt with in more detail. On the other hand the process of fluorescence, which is excited upon radiation in the characteristic absorption bands of coloured ions, exhibits the character of a monomolecular reaction.

The colour of the emission

In the case just discussed of the emission with the collaboration of foreign ions it will be clear that for non coloured ions (level E in fig. 13 *b* and *c*) it must be expected that the frequency of the light emitted will depend very much upon the height of the empty band above the occupied level E of the ion. Changes in the position of this band (for instance upon the replacement of the basic material by a similar material of altered composition) have a pronounced effect on the colour of the light emitted. A beautiful example of this is found in the mixed crystals of ZnS and CdS, activated with Cu or Ag. In the case of these crystals, with increasing percentage of Cd the limit of crystal absorption (λ_g) is shifted toward longer wave lengths. As may be seen from the table the maxima of the emission (λ_m) are shifted in the same direction with Cu as well as with Ag as activator.

In the case of the coloured ions, on the basis of the above, no effect on the colour of the emission may be expected by a change in the basic material. The emission proceeds by means of transitions between levels which correspond to more or less undisturbed movements of the electrons within the same ion.

atom per cent ZnS	atom per cent CdS	λ_g Å	λ_m(Ag) Å	λ_m(Cu) Å
100	0	3 360	4 600	5 230
85	15	3 640	4 920	5 790
68	32	3 920	5 410	
50	50	4 270	6 000	
33	67	4 700	6 550	

Colour changes have, however, also been observed here, which probably occur in a different way. In the first place it must be noted that in a substance like willemite the absorption bands of the Mn^{2+} ion lie partly in the ultra violet, while the emitted light is green. This can be interpreted in the following way (fig. 13*d*): primarily a coloured ion (energy level C) is brought into one of the highest excited states by the absorption of light. From there, *via* intermediate levels, it passes to the lowest excited level from which the electron returns to the basic state with the emission of radiation.

The transition from one level to the next lower level can be considered to occur directly, the excess energy being given off to the crystal lattice in the form of heat, or *via* processes of repeated emission and absorption. The latter is possible because of the fact that here also it may be expected that as a result of the collaboration of the crystal lattice the emission line which corresponds to a given absorption transition is shifted slightly with respect to the absorption line toward long wave lengths, and may therefore just coincide with a following absorption line or narrow absorption band. No definite statements can be made about this.

When there is more than one state from which return with emission is finally possible, different competing emission bands occur in the emission, which, when they have a certain width, as is usually the case, may overlap and apparently fuse into one band. If then, due to the influence of the surroundings on the electron movement in the ion upon change in the basic material, a change occurs in the chances of transition for the two emitting levels, this will cause a change in the distribution of the energy over the two emission bands, with the result of a shift of the maximum and an alteration in the colour. Examples of this mode of behaviour are the following. The dependence of the colour of the emission in the case of ZnS-MnS and (ZnMn)-silicate as a function of the manganese content and of (ZnBeMn)-silicate as a function of the beryllium content.

Phosphorescence

The concept of the energy bands already sketched also furnishes a useful explanation of the phenomenon of phosphorescence. In this case unoccupied disturbed levels (*F* in *fig. 14*) play a part. It is

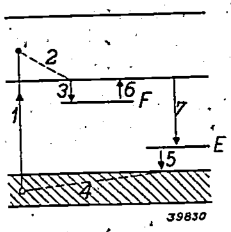


Fig. 14. Mechanism of phosphorescence.

possible that the electron from the empty band reaches these (fig. 14, *3*), while the open place in the empty band is filled by the electron from *E* (*5*). Return from *F* to *E* is then impossible if *E* and *F* do not lie very close to each other in space.

Return can then only take place by the electron first being brought back again from *F* to the empty band (*6*), which requires energy and is therefore only possible upon addition of energy, for instance from the store of heat of the crystal lattice. Return from *F* to the empty band can also take place under the influence of irradiation with infra-red or visible light.

When the electron is in the empty band, there is again opportunity of emission (*7*). At sufficiently low temperature and in the absence of external irradiation the state in which the electron is situated in *F* and in which therefore the crystal possesses a certain excess of energy in reserve may last for a considerable time.

Literature on luminescence

W. de Groot, Philips techn. Rev. 3, 125, 1938 (fluorescence).

J. H. Gisolf and W. de Groot, Philips techn. Rev. 3, 241, 1938 (fluorescence and phosphorescence).

A. A. Kruithof, Philips techn. Rev. 6, 65, 1941 (luminescence lamps).

P. Pringsheim, Fluoreszenz und Phosphoreszenz, Springer, Berlin 1928; Handbuch der Physik, vol. 23, part 1, Springer, Berlin 1933.

P. Lenard, Schmidt, R. Tomaschek, Handbuch der Experimentalphysik, vol. 23, part 1, Akad. Verl. Leipzig 1928.

F. Fritz, Leuchtfarben, Bodenbender, Berlin 1941.

N. Riehl, Lumineszenz, Springer, Berlin 1941.

F. A. Kröger, Luminescence in solids containing manganese (dissertation Amsterdam 1940).

SOME PARTICULARS OF DIRECTIONAL HEARING

by K. de BOER and A. TH. van URK.

534.76

While the perception of direction in hearing is explained by the time and intensity differences between the sound contributions which the two ears receive, other effects must be sought to explain the localization "in front" and "behind". Several effects which may here play a part are investigated with the help of an arrangement for stereophonic sound reproduction previously described in this periodical.

When listening to a source of sound which does not lie immediately in front of the head of the listener, but at an angle α to the vertical plane bisecting the head (*fig. 1*) the sound does not strike both ears at the same moment and the sound intensity is not the same for both ears. As was discussed in detail in an earlier article in this periodical [1]), the perception of direction which forms a part of hearing is mainly due to these time and intensity differences.

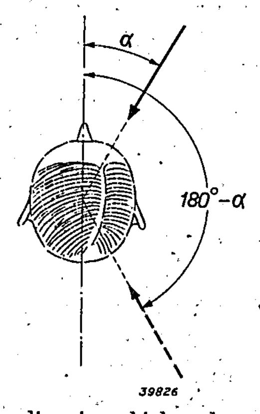


Fig. 1. Sound from a direction which makes an angle α with the direction in which the listener is looking reaches the two ears at different moments and (due to the refraction by the head) with different intensities. If we idealize the head to a sphere, the same time and intensity differences occur with the sound direction 180°—α as with the direction α.

By looking fig. 1 in which the shape of the head can be approximated by a sphere, it will be seen that the direction 180°—α is equivalent to the direction α as far as the time and intensity differences at the two ears are concerned. From this it would be concluded that the sense of hearing cannot distinguish between two such directions — one in front and one behind the head. Experience shows, however, that the sense of hearing can indeed localize a source of sound as "in front" or "behind". How can this be explained?

The simplest possible explanation starts from the fact that the human head actually deviates from the spherical form. In particular a dissymmetry is caused by the presence of the outer ears or auricles. Since the auricles have a shielding effect for sound waves of very short wave length, *i.e.* for high tones, sounds coming from the back must in general have a duller timbre than sounds from the front. This difference in timbre the hearer is assumed to interpret as a difference between front and back.

We have attempted to test this explanation with the help of an installation for stereophonic sound reproduction previously described in this periodical [2]). Use is here made of a spherical "artificial head" in which two microphones are placed at the ends of a horizontal diameter. Each microphone is connected *via* its own amplifier to one of two telephones over the ears of the listener. Artificial head and listener were now placed in different rooms, while in the vertical plane bisecting the artificial head a source of sound (speaker) was set up (*fig. 2*). Since the artificial head has no auricles, according to the hypothesis given it must be expected that in these tests the listener will be in doubt as to whether the sound comes from the front or the back.

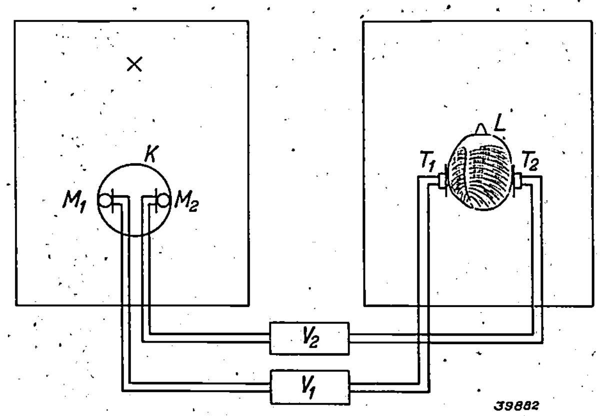


Fig. 2. Arrangement for stereophonic sound reproduction. In the vertical plane bisecting the artificial head K stands a speaker (×). Each of the two microphones M_1, M_2 of the artificial head is connected *via* a separate amplifier V_1, V_2 to one of the two telephones T_1, T_2 at the ears of the observer L who may be in a different room. With this arrangement the observer always hears the sound as if from behind.

---

1) K. de Boer, Stereophonic sound reproduction, Philips techn. Rev. 5, 107, 1940.

2) K. de Boer and R. Vermeulen, On improving defective hearing, Philips techn. Rev. 4, 316, 1939. In this article the problem of hearing from "in front" or from "behind" and also several of the experiments here described were already touched upon.

In our experiments this was by no means found to be true. All the persons used for the tests had an absolutely definite perception that the sound came from behind, whether the speaker stood in front of or behind the artificial head.

And still worse: when the sphere of the artificial head was replaced by a well modeled head with ears (large ones at that), the observer still continued unfailingly to hear the sound as coming from behind.

It is possible that this remarkable effect is caused by the fact that we are accustomed to observe a source of sound situated in front of us visually as as well as audibly, and thus upon missing the visual observation we choose the alternative of "behind us". A second possibility is that due to the limited frequency region in which the electro-acoustic apparatus used has a flat characteristic, an undesired auricle effect is introduced: the sound reproduced lacks the highest audible tones and may therefore make an impression of "dullness" on the listener which is interpreted by him as "sound from behind".

Whatever the truth of the matter may be, a decision as to whether the action of the auricles has a share in distinguishing between in front of and behind, cannot be reached in this way. Experiments with an apparatus whose frequency characteristic is flat up to the limit of audibility of about 16 000 c/s would perhaps furnish the desired answer.

In the meantime another explanation of the localization in front or behind also exists, namely that it is obtained by involuntary slight movements of the head of the listener [3]). In *fig. 3* the effect of such a movement is illustrated. If the source stands to the right of the middle, a turn of the head towards the right will bring the sound closer to the middle when it comes obliquely from the front (fig. 3*a*, *b*), and the same turn will take the sound farther away when it comes obliquely from behind (fig. 3*c*, *d*). The combination of two observations during the turning of the head thus provides an unambiguous conclusion as to the direction of the sound.

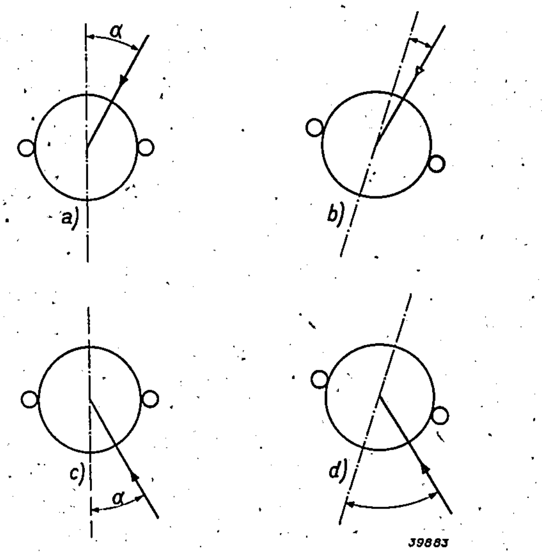


Fig. 3. While the two sound directions indicated in (*a*) and (*c*) cannot be distinguished by the observer, the difference is easily ascertained after a slight turn of the head (*b* and *d*).

This hypothesis can also be very simply tested with the arrangement described. It is exactly the advantage of the arrangement that the different effects which always occur simultaneously and more or less outside the control of the observer in ordinary hearing and whose influence cannot therefore be defined, can now be imitated separately. The effect of slight turns of the head is here investigated by allowing the artificial head and the head of the listener to turn back and forth around their vertical axes synchronously. The synchronism could be obtained in a very simple way by means of a lever in the hands of the speaker near the artificial head by means of which he can turn the artificial head slightly to the right and left in a rhythm which he indicates by counting out aloud. The listener seated in the other room hears the counting in his head phones and turns his head back and forth in the same rhythm. The result now is actually that the perception of "sound from in front" is obtained when the turning of the artificial head and listener's head take place in phase. If they turn in opposite phase, which amounts to the same thing as turning in phase in the case of a diametrically opposite positioning of the speaker with respect to the artificial head, the listener again hears the sound as coming from behind.

It is therefore quite admissible to suppose that in natural directional hearing also the listener ascertains the difference between front and back by small involuntary turns of the head. Such turns of the head occur for example automatically when the observer tries to see the source of sound, but also without this the head always makes slight, more or less reflex motions which may already furnish the desired effect.

In the experiments last described it is very tempting to try to confuse the hearing, namely by not moving the artificial head and the listener's head in the same way. The motion in "opposite phase" was already a sample of this, to which, however the hearing promptly furnished the correct answer. What will happen if we allow the artificial head

3) J. L. van Soest, Physica 9, 271, 1929; H. Wallach, J. Acoust. Soc. Amer. 10, 270, 1939.

and the listener's head to turn with different amplitudes? If we make the amplitude of the listener's head very small or equal to zero, while the artificial head turns with a large amplitude, a new phenomenon occurs: the listener simply hears the sound move back and forth according to the relative motion of the speaker with respect to the artificial head. If conversely we allow the head of the listener to turn back and forth while the artificial head moves less or not at all, a new impression is received: there is a perception of a definite direction of sound which no longer lies in the horizontal plane but possesses a certain angle of elevation.

This effect fits perfectly into the frame of the theory presented when it is kept in mind that a certain time and intensity difference of the sound at the two ears can actually be realized not only by one direction in front of (angle α) and one direction behind (angle $180°—\alpha$) as was sketched in the diagram on a plane in fig. 1, but by all directions which lie on a cone with the semi apex angle $90°—\alpha$ around the axis of the ear, see *fig. 4*. The choice among all these directions can be brought about in natural hearing with the help of the above mentioned slight turns of the head.

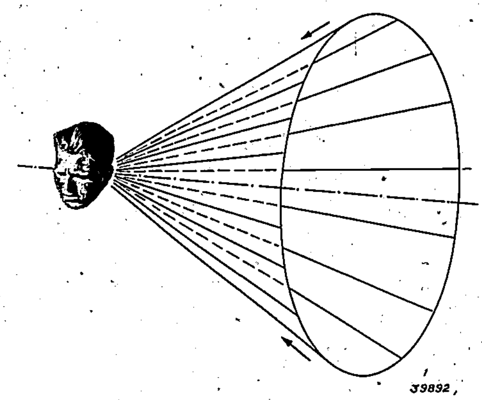


Fig. 4. Time and intensity differences are not equally great only for one direction forward and one direction backward, but for all directions which constitute the lines describing a cone about the axis of the ears of the observer.

To each position of the head corresponds a definite cone of directions, and the true direction of the source of sound must therefore be the common

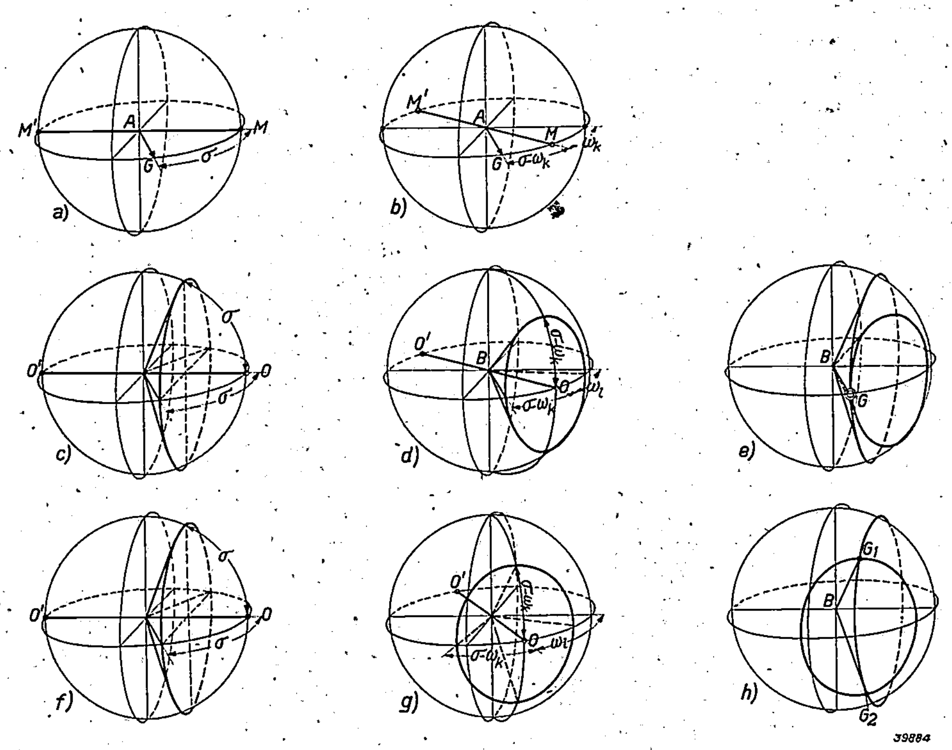


Fig. 5*a* en *b*). Axis of the ears MM' of the artificial head and direction AG of the source of sound before and after turning the artificial head through an angle ω_k.
c and *d*) Axis of the ears OO' of the listener and cone of the possible sound directions before and after the simultaneous turning of artificial head (angle ω_k) and listener's head (angle ω_l). In this case ω_l is assumed to be equal to ω_k. *e*). The two cones touch each other along the line BG.
f and *g*) The same as in *c* and *d*, but for the case where $\omega_l > \omega_k$. *h*) The two cones intersect each other in two lines BG_1 and BG_2, which deviate upwards and downwards, respectively, from the horizontal plane.

line of intersection of two or more such cones [4]).

This now also occurs in the experiments with the artificial head. The source of sound stands in the horizontal plane through the axis of the ears of the artificial head, in a direction which makes an angle σ with the axis of the ears (*fig. 5a*). The listener must choose the direction of the sound on a cone with the semi apex angle σ, see fig. 5*c* or *f*. If the artificial head now turns through an angle ω_k (fig. 5*b*) the angle between axis of ears and direction of sound becomes $\sigma-\omega_k$, the new cone of directions for the listener thus also takes on the semi apex angle $\sigma-\omega_k$; the axis of this cone has turned with the head of the listener through an angle ω_l. If $\omega_l = \omega_k$ (fig. 5*d*) as in the experiments initially described, the old and the new cone touch each other at a line in the horizontal plane which makes an angle σ with the old direction of the axis of the ears (fig. 5*e*). No angle of elevation is here observed and the sense of hearing takes a decision as to front or back only by means of the turn. If, however, $\omega_l > \omega_k$, see fig. 5*g*, then the old and the new cones clearly intersect each other in two lines which make a certain angle ϑ upward or downward with the horizontal plane (fig. 5*h*) [5]).

The angle of elevation which must be observed in carrying out the experiment with certain angles ω_k and ω_l can easily be calculated according to the rules of spherical trigonometry. In *fig. 6*

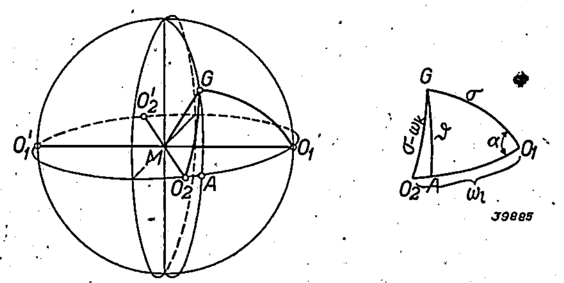


Fig. 6. The angle of elevation ϑ of the sound direction, which is observed according to fig. 5*h* upon turning the listener's head back and forth with greater amplitude than the movements of the artificial head, can be calculated with the help of the spherical triangles O_1AG and O_1O_2G.

O_1O_1' is the first position of the axis of the ears of the listener, O_2O_2' the second position, while MG represents the line of intersection of the two cones of observations (we consider only the one directed upwards). In the spherical triangle O_2O_1G the side $GO_1 = \sigma$, since G lies on the cone with semi apex angle σ and axis O_1O_1'. Further the side $GO_2 = \sigma-\omega_k$, since G also lies on the cone with semi apex angle $\sigma-\omega_k$ and axis O_2O_2'. Finally the side $O_2O_1 = \omega_l$, namely the angle through which the head of the listener was turned. For the angle α in the spherical triangle mentioned we now find

$$\cos \alpha = \frac{\cos(\sigma-\omega_k) - \cos\sigma \cos\omega_l}{\sin\sigma \sin\omega_l};$$

while the required angle of elevation ϑ of the direction MG can be calculated from the right-angled spherical triangle O_1AG, namely

$$\sin\vartheta = \sin\alpha \cdot \sin\sigma.$$

If one considers only the case in which the source of sound in the state of rest lies in the vertical plane bisecting the artificial head, so that $\sigma = 90°$, the formulae become simpler and one finds directly the relation between ω_k, ω_l and ϑ:

$$\cos\vartheta = \frac{\sin\omega_k}{\sin\omega_l} \quad \ldots\ldots \quad (1)$$

For the extreme case where $\omega_k = 0$ (artificial head stationary, only the listener's head turns back and forth) this whole method of representation is easily verified. The listener must always hear a source of sound which stands in the vertical plane bisecting the artificial head directly above or below himself, since $\cos\vartheta = 0$, $\vartheta = 90°$. This is indeed immediately clear, because the artificial head, and thus also the listener always hears the same thing in this experiment, and this is only possible when the listener's head turns back and forth around a vertical axis, at a position of the source of sound at the zenith or nadir. This observation was indeed confirmed in the experiments, with the expected uncertainty between "above" and "below" some observers always heard the sound from above, others from below, from their own abdomen.

For angles $\omega_k \neq 0$ equation (1) is not so easy to verify, since in order to do so the head of the listener and the artificial head must move synchronously each with an accurately determined amplitude. This could be realized with the arrangement represented in *fig. 7*. Artificial head and observer are in two adjacent rooms which are separated by a sound-proof glass wall. A pointer with a white disc is attached to the artificial head and this is watched by the observer. When (without turning his eyes) the observer follows the white disc as the artificial head turns, his own head turns synchro-

4) The "construction" in this way of an angle of elevation is usually quite difficult for the ear, it therefore prefers to use a kind of zero-point method: with a sound coming obliquely from above the head is involuntarily put into such an oblique position by trial and error that the direction of the sound is brought into the "horizontal" plane through the axis of the ears.

5) The decision between "above" and "below" can be obtained in natural hearing by turning the head about a different axis.

nously with the artificial head and there is a definite relation between the two amplitudes. This relation can be varied very simply by changing

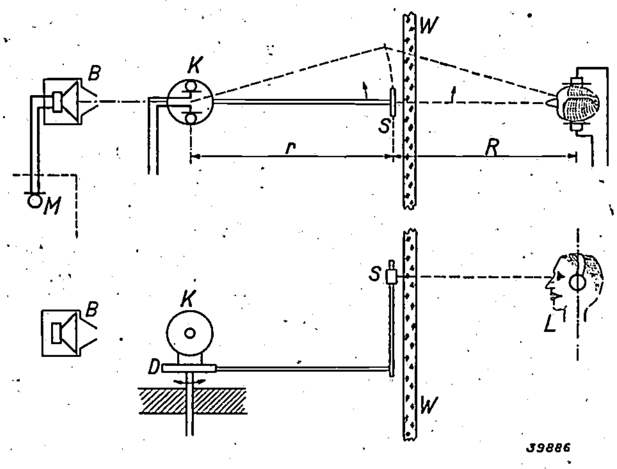


Fig. 7. Arrangement for the experimental determination of the observed angle of elevation when the angles ω_k and ω_l are different. K artificial head, L listener's head, W soundproof glass wall. B source of sound (loud speaker with very small opening connected with a microphone M set up in another room, before which stands a speaker), D turntable on which the artificial head is mounted and to which is fastened a lath with a white disc S as pointer.

the length of the pointer (r) or the distance between observer and glass wall (R). In order to fix satisfactorily the relative position between source of sound and artificial head, a loud speaker with a very small opening is placed in front of the artificial head. The fact that the artificial head and the listener's head here turn in opposite directions can of course be corrected by a suitable connection of the two head phones of the observer. The white disc is adjustable in height in order to be sure that the head of every observer actually turns about a vertical axis.

With this arrangement the perception of every angle of elevation between 0 and 90° could be obtained. A certain difficulty occurred, however, in measuring the angles of elevation observed. The observer could of course indicate with outstretched arm the direction perceived, but this was not found to be a suitable method for measuring the angle. Therefore use was also made of the fact that together with the perception of direction there is also a certain perception of distance which depends upon the intensity of the sound. With a certain intensity of the sound the observer has the feeling that the source of sound is situated on the outside of his head, whereupon he can indicate the spot (the sound image) with his finger [6]). In this way the angles of elevation observed (angle between the horizontal plane and the perpendicular from the sound image to the axis of the ears) could be measured with sufficient accuracy.

In *fig. 8* we have plotted the calculated and measured angles of elevation ϑ as functions of each other [7]). Theoretically all the points should lie on the 45° line. The agreement is as good as could be expected in these difficult measurements.

If we consider the experiments described as attempts to confuse the ear, it is clear that the sense of hearing possesses sufficient adaptability to find an unambiguous solution in all kinds of peculiar situations. It sometimes even registers perceptions which it certainly could not have learned from experience, but which are as it were extrapolated. Thus the localization of sounds inside the body — especially in the head, see footnote [4]) — or the perception of direction previously discussed (see the article cited in footnote [1])), which is caused by differences in time which are many times greater

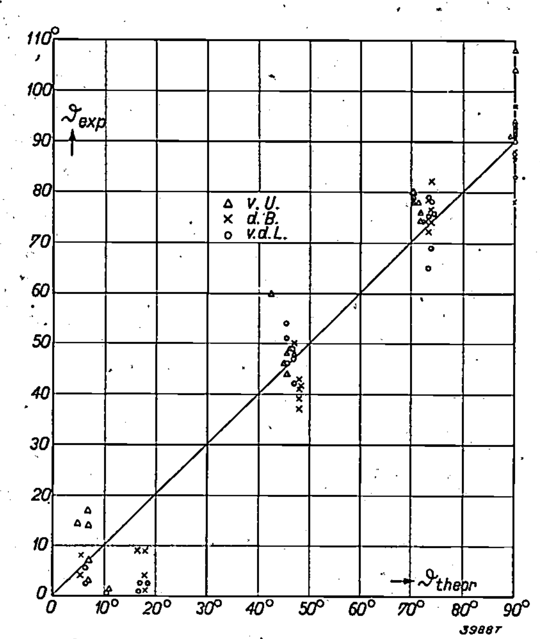


Fig. 8. The elevation angles $\vartheta_{exp.}$ with different turns of artificial head and listener's head plotted against the angles calculated $\vartheta_{theor.}$. The triangles, circles and crosses refer to three different observers. The points measured show great scattering but no systematic deviation from the theoretical 45° line.

---

6) It is then remarkable that one observer hears a more intense source of sound farther away, while another still "closer", *i.e.* in the head! For a weaker source of sound the reverse is then true.

7) In the calculation of ϑ a correction must still be introduced because of the fact that the heads of the observers compared with each other and with the artificial head do not have the same diameter. In our case the artificial head had a circumference a_k, which was smaller than the circumference (a_l) of all the observers' heads. Upon a turn ω_k of the artificial head therefore the time and intensity differences at the microphones are too small. This was taken into account roughly by inserting in equation (1) instead of the true angle ω_k, a smaller angle $\omega_k \cdot a_k/a_l$.

than the maximum differences in time of arrival at the two ears which can occur in normal hearing.

Nevertheless the sharpness of hearing also has its limits, and these are apparently exceeded when in the experiments for the measurement of the angles of elevation observed we exchange the two head phones of the listener. With equal turns back and forth of artificial head and listener's head, $\omega_k = \omega_l$, as was to be expected, the perception occurs that the sound comes from behind. With unequal turns, on the other hand, with $\omega_l > \omega_k$, whereupon the ear should register an angle of elevation towards the rear, the perception becomes very vague and difficult to define. In *fig. 9* a comparison is given between the elevations observed and those calculated for this case. The spreading is greater than in fig. 8, the measured points with $\vartheta_{\text{theor.}} \approx 65°$ all lie too low, some measurements here are entirely wrong. It is remarkable that in these measurements much more clearly than in the measurements in front, the theoretically expected indefiniteness occurred between directions above and below the horizontal plane. Of the three observers whose measurements are reproduced in fig. 9 one always heard the sound above, one always below and one could localize it at will either above or below the horizontal plane, while in the measurements of fig. 8 the localization was quite regularly above the horizontal plane.

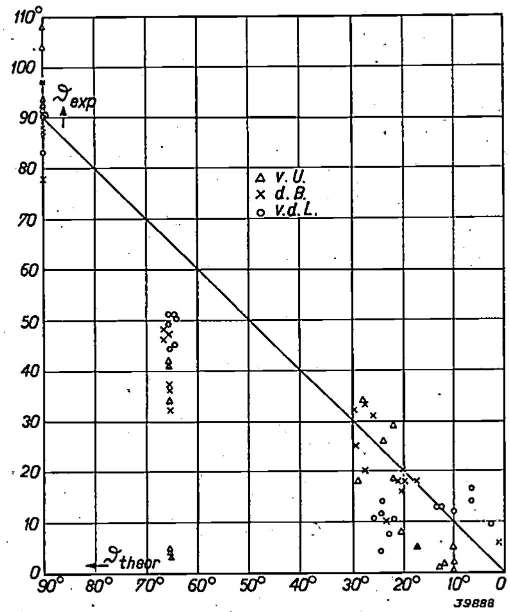


Fig. 9. As in fig. 8, but for sound directions from behind. The measured points for $\vartheta_{\text{theor.}} = 90°$ (artificial head stationary) are of course the same as in fig. 8.

THE PERMEABILITY OF METAL WALLS FOR GASES

by J. D. FAST.

620.193.29

A survey is given in this article of the phenomena which play a part in the permeation of gases through metal walls. It is pointed out that the gases in the metal are present in atomic (or perhaps ionic) form, and that they move in the space between the atoms of the crystal lattice without this movement being necessarily accompanied by a diffusion of the metal itself. Attention is also called to the part played by the processes which take place at the surface of the metal, namely the dissociation into atoms and the recombination to molecules, respectively, and the penetration into and the emergence from the metal, respectively, of the atoms.

Introduction

In different types of high vacuum tubes the walls are made wholly or partly of metal. In some cases this type of construction is employed in order to obtain effective cooling by means of running water, in other cases in order to ensure a simplified method of mass production.

The problem of mass production is particularly important in the case of radio valves where in different cases the glass walls have been replaced by metal.

The cooling problem is of particular importance in transmitter valves, rectifier valves, X-ray tubes and the like, in which in every case only a part of the energy supplied can be converted into the desired form and where the dissipation of the lost heat energy in high-power tubes forms one of the most important problems.

Although in such cases metal walls offer advantages over glass walls, difficulties may also occur in their use which are almost entirely absent when glass is used. One of the most important of these is the fact that upon contact of the outside surface with cooling water as well as with gases the vacuum on the inside may under certain circumstances deteriorate, due to the fact that gases diffuse to the inside through the walls. In this article we shall try to explain the ideas which may be gained of this transmission of gases through metal walls. It will be found that the phenomenon is more complicated than is expressed by the term diffusion. In addition to this latter process, which takes place inside the metal, we are also concerned with processes which take place on the surfaces of entrance and emergence. From the discussion the relation will also appear which exists between the transmission and other more familiar phenomena which occur upon contact between gases and metals, such as adsorption and solution.

The mechanism of diffusion in metals

Geometric and chemical factors

At first glance it will seem strange that gases should be able to move through compact metals, and the first question which arises is about the mechanism of this process. It should be stated in advance that very little is yet known with certainty about it. In the attempt to form an idea of the different possibilities we shall not therefore limit ourselves to the diffusion of gases, but shall treat the phenomenon of the transport of matter in solid metals more in general.

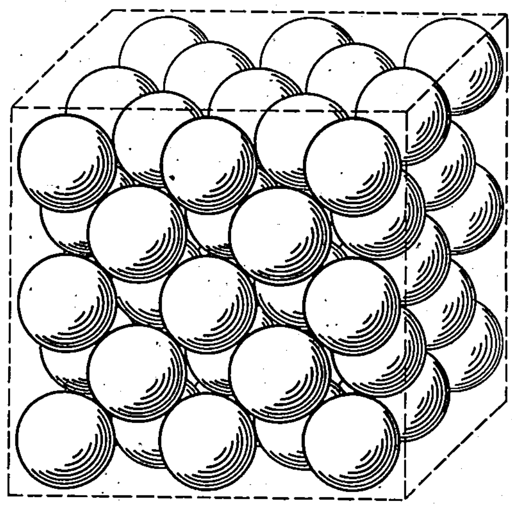


Fig. 1. Closest cubic packing. The lattice is face-centred. Examples of metals which crystallize in this way are: copper, silver, aluminium, nickel, palladium, platinum and furthermore carbon-free iron above 900 °C.

A primitive idea can be obtained of the structure of a metal crystal by considerng it to be built up of atoms in the form of small hard spheres with a radius of about 10^{-8} cm, which are packed together in such a way that a regular three-dimensional lattice results. In the case of most metals the symmetry of the lattice is such that the spheres fill the available space as closely as possible. This can be attained by two different ways of packing and leads to the familiar structures with cubic and hexagonal closest packing. In both cases 74 per cent of the total volume is occupied by the spheres. In *figs. 1* and *2* the arrangement of the atoms for these two closest packings are represented. Another group of metals has a slightly less close packing

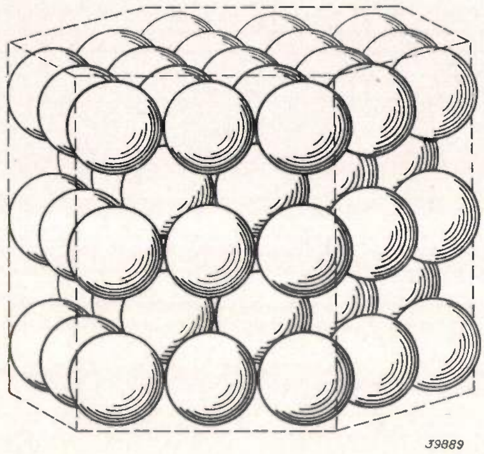


Fig. 2. Closest hexagonal packing. Examples of metals which crystallize in this way are: beryllium, magnesium, titanium and zirconium, the last two below their respective transition points at 885 and 865 °C.

with cubic body-centred structure, in which the spheres occupy 68 per cent of the available volume (see *fig. 3*).

With the picture of a metal crystal just given it is difficult to imagine the occurrence of diffusion, unless it be that of foreign atoms (likewise considered to be spherical) which are very small compared with the metal atoms. Foreign atoms can only move unhindered through the sphere arrangements described when they are even smaller than the largest spheres which could be placed in the cavities between the large packed spheres.

The image of a metal crystal as a symmetrical arrangement of hard spheres is of course much too primitive. The idea which is now held of the structure of a metal is that of a three-dimensional lattice of positive ions which (floating freely as it were) are situated in a dense electron "gas", which gas uniformly fills the spaces between the ions. Due to mutual electrostatic repulsion the ions tend to remain as far away from each other as possible. The mutual distances are here determined in such a way that the pressure of the electron gas and the electrostatic attraction between ions and electrons maintain an equilibrium which at the same time results in the arrangement in a regular lattice. The image of the symmetrically arranged spheres can now if desired be transferred to the ions, but the spheres must now no longer be considered as entirely hard, while moreover they need not touch each other [1]). In the case of the alkali metals the ions lie at relatively great distances apart, in the case of many other metals they lie closer together. In *fig. 4* the position of the ions in the electron gas for the closest packed lattice planes is shown for the metals sodium and copper [2]). In sodium about 70 per cent of the volume of the metal is occupied by the electron gas. In copper, a subordinate series metal, on the other hand, the ions are so close to each other that a certain mutual penetration of their electron shells even occurs.

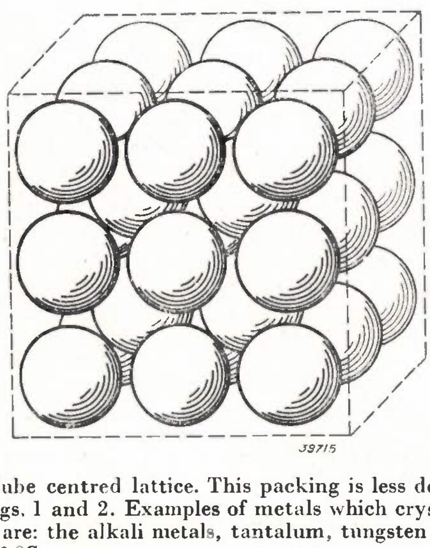


Fig. 3. Cube centred lattice. This packing is less dense than that in figs. 1 and 2. Examples of metals which crystallize in this way are: the alkali metals, tantalum, tungsten and iron below 900 °C.

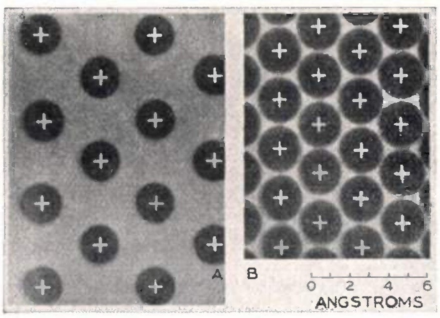


Fig. 4. Arrangement and true relative size of the ions in the most densely occupied lattice planes of sodium and copper according to Shockley. The boundaries of the ions are idealized, actually the density of electrons decreases gradually from a high value inside the ion to a much smaller constant value outside of it.

The spaces between the atoms may therefore in many cases be larger than would be expected on the basis of the primitive picture. On the other hand, these spaces are not empty, as in the primitive con-

[1]) If the modern idea of an atom (or ion) as a positive nucleus surrounded by a cloud of electrons is kept in mind, the concept of "touching" already becomes very vague.

[2]) The figure is borrowed from an article by W. Shockley, J. Appl. Phys. **10**, 543, 1939. References to the literature are here made only when no reference was made to it in a publication which will shortly appear in "Physica" under the same title. For further literature references therefore the reader should consult that article.

cept, but filled with electrons. A diffusion of a gas through a metal is therefore scarcely conceivable without considerable reciprocal action between the two substances. It may therefore be expected that the diffusion of gases through metals will in general be of a specific character. It is found indeed that a gas is only able to dissolve in a metal and diffuse through it when there is a certain "chemical affinity" between the two. Thus for instance oxygen can be taken up by copper and silver and can also diffuse in these metals, nitrogen on the other hand cannot do this, while chemically also it has little tendency to combine with these metals. The latter gas can, however, at sufficiently high temperatures, diffuse through iron and alloys of iron and chromium, while compounds of these metals with nitrogen are known. The rare gases, which form no chemical compounds, are insoluble in all metals and diffuse through none of them.

Influence of temperature

In order to understand the occurrence of diffusion in a solid metal wall, it must also be kept in mind that not only do the foreign atoms execute a heat motion, which makes the diffusion possible, but that the lattice ions also do so in the form of vibrational movements around their positions of equilibrium. With increasing temperature this causes at the same time an expansion of the whole lattice. The presence of the gas itself causes in addition an extra deformation (expansion) of the lattice. All these factors create in many cases a possibility, even for relatively large atoms, of moving by jumps through the metal lattice from interstice to interstice at temperatures which are not too low. In doing this they must as it were push through between the ions of the lattice at suitable moments, thereby overcoming repulsive forces. Thus only those atoms can make a "jump" which possess at least a definite energy E, which will usually be large compared with the average energy of the atoms. That such atoms which are rich in energy are also present at thermal equilibrium is due to the fact that according to the Maxwell-Boltzmann law of distribution the energies extend to very high values. The fraction of the atoms which have at least a certain high energy E is given by:

$$\frac{\Delta N}{N} = e^{-E/kT} = e^{-Q/RT} \quad . \quad . \quad . \quad (1)$$

(E = energy per atom, k = Boltzmann's constant, Q = energy per gram atom, R = the gas constant). From this it follows that the degree to which the "jumps" occur is determined in the first place by the temperature, and that the diffusion constant δ will be approximately proportional to the factor given by equation (1) in which the name heat of diffusion can be given to the "activation energy" Q.

We thus find that

$$\delta = Ae^{-Q/RT}, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where A is a constant.

Solution and diffusion by substitution

The form described of diffusion of atoms of an element B through the interstices of the lattice of an element A will be expected especially when the atomic volume of B is relatively small and the atoms of B can be easily deformed (for instance elements with a low atomic number such as hydrogen, boron, carbon, nitrogen and oxygen), or when the lattice of A has especially large interstices. This expectation is naturally coupled with the assumption that the substance A permits the presence of atoms of B in its lattice interstices, that is to say that A possesses a certain solvent capacity for B. This, as was mentioned in the preceding, is determined not only by geometric but also by chemical factors [3]).

In addition to the form of solution and diffusion considered until now, however, the case often occurs in which the foreign atoms B occupy positions in the lattice which in the pure metal A are occupied by atoms of A. This is the case, in particular, upon the solution of a foreign metal in another metal. Diffusion phenomena may also occur in such substitutional solid solutions under suitable conditions, whereby any possible differences in concentration are equalized. Different conceptions may be formed of the mechanism of diffusion in substitutional solid solutions. It may for example be assumed that at a sufficiently high temperature a foreign ion B changes place here and there in the lattice with an adjacent ion A, so that two ions are concerned in each "jump". It is, however, more probable that only one ion is concerned in each "jump". It must then be assumed that at high temperatures in substitutional solid solutions also a small part of the atoms are present in the interstices of the lattice, or that there are unoccu-

[3]) Strictly speaking it is not permissible to make a sharp distinction between chemical and geometric factors, since all the forces active in the binding, even the repulsive forces, with which any spatial limitations occurring are closely connected, must be considered as belonging to the "chemical" forces.

pied positions in the lattice. In the first case the diffusion occurs in the way already discussed, namely by "jumps" from interstice to interstice. In the second case an ion can jump to an adjacent unoccupied position, and it is clear that a diffusion current of ions in a given direction according to this mechanism is equivalent to a current of unoccupied positions in the opposite direction. In considering fig. 4 both types of "defects", namely unoccupied positions and ions in the interstices can be imagined in a metal like sodium. In a metal like copper on the other hand one can in the first instance imagine the presence of unoccupied places only.

In the so-called theory of disturbed order of Wagner and Schottky [4]) it is assumed that at a temperature above the absolute zero point a crystal, in the thermodynamical equilibrium state always contains a number of the "defects" spoken of. If these reversible defects of atomic dimensions are held responsible for the diffusion, the foreign atoms B and the lattice atoms A move independently of each other in the first instance. However, the movements are correlated with each other in as much as the current of foreign atoms B in the one direction cannot be larger than the maximum possible current of atoms A in the other direction. If it were otherwise the character of substitutional solid solution would be lost. The velocity of transport of B is limited therefore by the velocity of self-diffusion of A, which is in general small.

Hydrogen-iron and oxygen-zirconium

Making use of the conclusion just deduced, it can immediately be shown for different solutions of gases in metals that one is certainly not concerned with substitutional solid solutions. Thus for example hydrogen in metals like palladium and iron already exhibits high diffusion velocities at temperatures below 100 °C, although there is certainly no question of an appreciable self-diffusion of the pure metals or of the metals containing hydrogen. It might now be assumed that in this case the movement of the hydrogen takes place along the crystal boundaries. The metals employed technically never consist of a single crystal, but of many crystals grown together. These crystals exhibit great relative differences in orientation, and it is clear that the packing of the atoms (ions) at the boundaries will be less perfect than in the interior of the crystals. Cases are indeed known in which the diffusion takes place primarily along these boundaries. A familiar example is the diffusion of thorium through tungsten [5]). For the system iron- hydrogen, however, it could be shown that the permeability of a single crystal is the same as that of a polycristal, so that the movement undoubtedly here takes place through the lattice.

In other cases there is a possibility of demonstrating the occurrence of the interstitial type of solution from a combination of X-ray measurements of the lattice constants and direct measurements of the specific weight. Thus for example the specific weight of zirconium is found to increase upon the absorption of oxygen, although oxygen atoms are much lighter than zirconium atoms. This already indicates an occupation of the interstices and not a substitution. unless a contraction of the lattice due to the presence of oxygen is assumed. In an X-ray examination, however, the distances between the zirconium atoms were found to have become greater. There is even a quantitative agreement between the directly measured (with a hydrostatic balance) specific weight and the value calculated from the lattice constants and the composition, if in the calculation it is assumed that all the oxygen is taken up in the interstices and no zirconium atoms in the lattice are replaced by oxygen.

Until now no cases are known in which gases form substitutional solid solutions with metals. In agreement with these gases, as far as they are soluble, exhibit in general high diffusion velocities in metals.

State of charge of the dissolved gas

From the occurrence of the interstitial type of solution and diffusion, for geometrical reasons it may be prophesied with fairly great certainty, as was tacitly assumed above, that the gases will be present in the metal, not in the form of biatomic molecules (H_2, O_2, N_2, etc.) but split into atoms. In agreement with this it is found in solubility experiments that the concentrations in metals of the gases mentioned increase at constant temperature with the square root of the pressure of the gas outside the metal.

By making use of modern conceptions about the metallic state we arrived in the foregoing at a picture of diffusion in pure metals and alloys in which it were not neutral atoms but ions which diffuse. It will further be expected that gases also in the medium of metal ions and electrons will not be present in the form of neutral atoms, but will in general

[4]) C. Wagner and W. Schottky, Z. phys. Chem. B 11, 163, 1930.

[5]) P. Clausing, Physica 7, 193, 1927.

carry an electric charge. It is found indeed in the few cases in which this phenomenon has been sought that electrolysis effects can occur in solutions of gases in solid metals. Under suitable conditions of temperature and field strength hydrogen in palladium moves toward the cathode, while nitrogen in iron and oxygen in zirconium move toward the anode. Because of the fact that these gases are present in the interstices of the lattice and have a relatively high mobility there, it may be concluded with a fair degree of certainty that hydrogen has a positive charge and nitrogen and oxygen negative charges. This does not mean that clearly distinguishable H^+ and O^{--} ions must be assumed, but rather that the distribution of the electron gas is such that considering an average over a longer time the hydrogen atoms carry a deficit and the oxygen atoms an excess of negative charge.

If we were concerned with substitutional solid solutions the conclusion about the sign of the charge would certainly not have been permissible. Let us for example assume that the motion in this case takes place *via* that type of reversible "defects" in which a very small fraction of the atoms are situated in the interstices. A and B then move independently of each other in the first instance and the concentration at the cathode will increase for that component for which the product of concentration C in the interstices, charge z and diffusion constant D has the largest positive value [6]). The magnitude of the ionic current is determined not only by the number of mobile charge carriers, but also by their velocity, and in an electric field of a given intensity the latter is proportional to the product zD.

If on the other hand it were assumed that in the substitutional solid solutions mentioned the motion takes place by direct exchange of place of approximately equally large A and B ions, it would be expected that the concentration of the ions with the largest positive charge will increase at the cathode, since the potential energy thereby decreases. If A and B do not have about the same atomic volume, so that a change in the lattice constant occurs due to the changes in concentration, then for the same reason it will be expected that the direction of motion will be determined not by the charge, but by the density of charge, *i.e.* by the quotient z/v (v = atomic volume). Schwarz [7]) has shown that this is the case with liquid alloys (amalgams).

If the hydrogen is actually present in the metals more or less as positive ion, unusually high diffusion rates may be expected for this gas. A positive hydrogen ion is nothing else but a hydrogen nucleus, a proton, whose dimensions are negligibly small compared with those of an atom (the radius of an atom is of the order of magnitude of 10^{-8} cm, that of a proton about 10^{-13} cm). Even at room temperature indeed hydrogen is found to be able to move through different metals, The velocities are, however, smaller than would be expected in case of a complete dissociation into protons and electrons. If, conversely, at a relatively low temperature an unknown gas permeates through a metal wall, the first thought will be that it is hydrogen.

For the sake of completeness we wish to point out that it is not impossible that the dissolved hydrogen ions bear a negative charge in certain metals, particularly the alkali and alkaline earth metals. We shall consider sodium as an example. An investigation by Hüttig and Brodkorb [8]) has shown that it is probable that this metal can contain 3 to 5 atomic per cent of hydrogen in solid solution, and that only after this amount has been exceeded is the hydride NaH formed. In this compound the hydrogen more or less plays the part of a halogen, and it seems obvious to assume that the dissolved hydrogen atoms also carry a negative charge and in electrolysis experiments will move toward the anode, but with much lower velocity than that with which in iron or palladium for instance they move toward the cathode.

The part played by the boundary surfaces when gases pass through metal walls

In the preceding we have spoken only of the movement of gases in the interior of solid metals. Since the gases diffuse in the form of atoms (or ions) it is obvious to assume that they can only penetrate a metal wall after a splitting of the molecules into atoms (or ions) has taken place on the entrance surface. On the emergence surface the reverse process (the recombination to molecules) must take place before the gas can leave the metal. In order to make the description of the permeation of gases through metal walls complete, therefore, in addition to the diffusion in the metal we must also consider the phenomena which take place in the adsorbed layer.

Two different types of adsorption of gases on metals can be distinguished. At a low temperature (for instance at the temperature of liquid air) so-called physical adsorption takes place, in which the gas is bound in a molecular form by van der Waals forces, *i.e.* by forces which are of the same nature as the forces of attraction acting between the molecules of gases, which are responsible for the deviations from the equation of state of the ideal gas and for condensation at low temperatures. This

[6]) C. Wagner, Z. phys. Chem. B **15**, 347, 1932; Z. phys. Chem. A **164**, 231, 1933.

[7]) K. E. Schwarz, Elektrolytische Wanderung in flüssigen und festen Metallen, J. A. Barth, Leipzig 1940.

[8]) G. F. Hüttig and F. Brödkorb, Z. anorg. Chem. **161**, 353, 1927.

adsorption decreases with increasing temperature. At higher temperatures so-called activated adsorption takes place in which chiefly chemical forces are active [9]). This type of adsorption is distinguished from the van der Waals adsorption by a much higher heat of adsorption, the existence of an activation energy and a more specific character. It is accompanied by dissociation of the gas into atoms and thus forms the required link in the process of permeation.

The fact that activated adsorption is actually accompanied by a splitting into atoms (thus for instance in the case of H_2 by the formation of a surface hydride) is illustrated in perhaps the most convincing way by experiments on the simultaneous adsorption on metals in powder form of ordinary hydrogen H_2 and the isotope of hydrogen D_2 with double the atomic weight (deuterium). In the temperature region of van der Waals adsorption the molecules of H_2 and D_2 are freed as such upon desorption. In the temperature region in which activated adsorption occurs a partial conversion into HD must be expected. If we assume for example that an H_2 and a D_2 molecule are adsorbed side by side on the metal surface according to the schematic representation below (*fig. 5*) it will be seen that upon desorption there is approximately just as great a chance that the gas will be freed in the form of $H_2 + D_2$ as in the form of 2 HD [10]). The results of experiments are in agreement with this conception [11]).

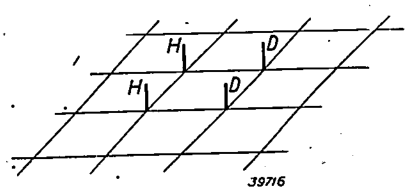


Fig. 5. Atomic adsorption of $H_2 = H + H$ and $D_2 = D + D$ on a metal surface. Possibility of the formation of HD by recombination.

The phenomena occurring upon adsorption and desorption are represented very comprehensively by a scheme of potential curves [12]). Curve *1* of *fig. 6* relates to the adsorption of a hydrogen molecule; curve *2* to that of two hydrogen atoms. The splitting of the molecule into atoms at a large distance from the surface requires the dissociation

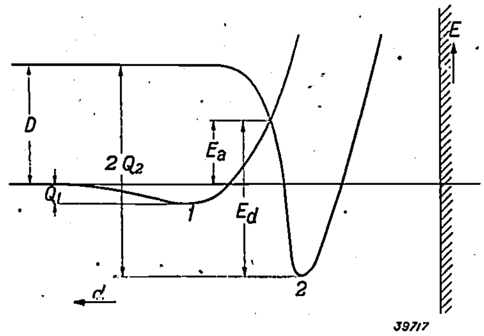


Fig. 6. Potential curves for van der Waals adsorption (1) and atomic adsorption (2) of hydrogen. Q_1 = molecular adsorption energy, Q_2 = atomic adsorption energy ($2Q_2$ idem for the adsorption of two atoms). E_a is the activation energy of the atomic adsorption; E_d idem for desorption (recombination), D is the dissociation energy of the free molecule.

energy D. Q_1 is the heat liberated in physical adsorption, $2Q_2$—D that liberated in activated adsorption (Q_2 is here the adsorption energy of one hydrogen atom, taken strictly at the absolute zero point and neglecting the zero point vibration). The height E_a of the intersection of the two curves above the zero line gives approximately the activation energy of activated adsorption. E_d indicates approximately the activation energy of desorption.

In the cases in which the gas atoms in the dissolved state carry an electric charge a binding in the form of neutral atoms will not be expected in activated adsorption. In the case of hydrogen for example, in the activated adsorption of each H-atom the electron will already be more or less taken up by the metal. In agreement with this it is found that the adsorption of hydrogen by many metals causes a lowering of the work function of the electrons. Conversely, the adsorption of oxygen causes an increase in the work function [13]).

Upon the passage of a gas through a metal wall, after the formation of atoms (or ions) on the surface, a new activation energy must be supplied to bring the atoms from this (chemisorbed) state into the interior of the metal. In the metal itself, for every jump from interstice to interstice, the activation energy of diffusion must be supplied. Upon reaching the surface of emergence an activation energy must again be supplied to cause the atoms to pass from the dissolved to the adsorbed state. As

[9]) H. S. Taylor, J. Amer. Chem. Soc. 53, 578, 1931.

[10]) In the temperature region concerned with permeability one must not, as is suggested by the figure, expect a binding of the H and D-atoms at fixed positions, but a state in which a large fraction is always moving from point to point over the surface by means of "jump"-like movements. This, however, makes no difference as far as the result to be expected is concerned.

[11]) M. Polanyi, Sci. J. Roy. Coll. Sci. 7, 21, 1937; A. Farkas, Orthohydrogen, Parahydrogen and Heavy Hydrogen, Cambridge University Press, Cambridge 1935.

[12]) J. E. Lennard Jones, Trans. Faraday Soc., 28, 341, 1932.

[13]) *Cf.* J. H. de Boer, Electron Emission and Adsorption Phenomena, Cambridge University Press, Cambridge 1935.

final step in the process comes the recombination of the atoms on the emergence surface and the desorption of the molecules formed, for which the activation energy E_d (fig. 6) is required.

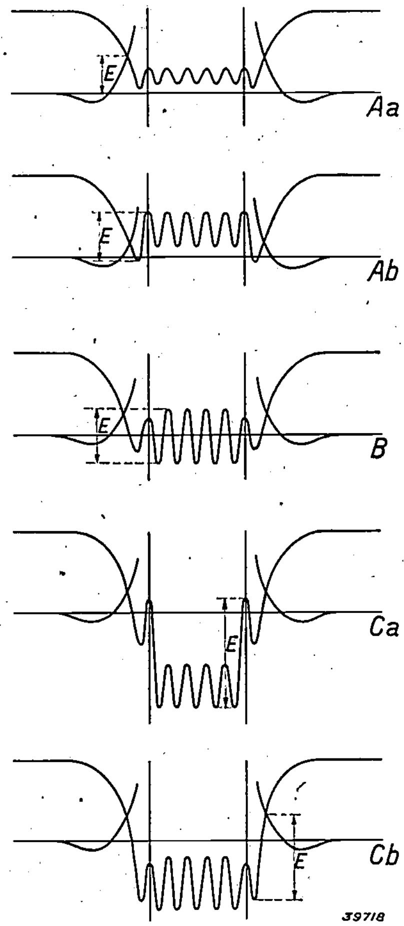


Fig. 7. Schematic representation by means of potential curves of the permeation of a gas through a thin metal wall. In *Aa* the dissociation of molecules into atoms on the entrance surface requires the greatest activation energy, in *Ab* the transition of the adsorbed atoms to the interior of the metal, in *B* the diffusion in the metal, in *Ca* the transition of the dissolved atoms to the emergence surface, in *Cb* the recombination of the adsorbed atoms on the emergence surface.

In principle of course it is possible to make the metal walls so thick that the diffusion process proper is the slowest of the various processes which occur in the permeation of gas. With not too thick walls, such as are used in the experimental investigation and in technical problems, the following theoretical possibilities can be distinguished.

A) The rate of permeation is mainly determined by the reaction at the entrance surface. Two cases are then possible:
   a) The rate determining process is the transition from van der Waals adsorption to activated adsorption, *i.e.* the dissociation of molecules into atoms (or ions) on the surface.
   b) The rate determining process is the transition of the atoms from the chemisorbed state to the inside of the entrance surface.

B) The rate of permeation is mainly determined by the diffusion in the interior of the metal.

C) The rate of permeation is mainly determined by the reaction at the emergence surface. There are again two possibilities:
   a) The rate determining process is the transition of the atoms from the inside of the emergence surface to the outside of this surface, *i.e.* the transition from the occluded to the chemisorbed state.
   b) The velocity determining process is the recombination of the adsorbed atoms.

These different possibilities are represented schematically by the composite potential curves of *fig.* 7. From the known facts it appears that all five types can be realized with different combinations of gas and metal and different conditions of the boundary surfaces. Examples of these five cases in order are the following:

Aa) hydrogen through an iron wall;
Ab) hydrogen through an iron wall with a very rough surface;
B) hydrogen through a copper wall;
Ca) oxygen through a zirconium wall;
Cb) oxygen through a copper or nickel wall.

In a later article in this periodical we hope to go into these and other examples of the permeation of gases through metal walls and the practical conclusions which may be drawn from them.

THE DETERMINATION OF THE ELASTIC CONSTANTS OF METALS

by M. J. DRUYVESTEYN.

620.172.225 : 621.317.755

The elastic constants of materials can be calculated from the characteristic frequencies of test rods in longitudinal and torsional vibration. In this article an arrangement is described with which these frequencies can be measured and with which measurements of the elastic constants of non-ferromagnetic metals were carried out. With the values of the constants found it is possible to check the isotropy of the substance. In certain cases large errors can be made if this check is omitted. This is explained in connection with the case of β-brass.

The modulus of elasticity E of a rod can be found by determining the elongation of the rod upon the extension of a tensile force P. E is given by the formula:

$$E = \frac{P}{d}\,\frac{l}{\Delta l}, \qquad (1)$$

where d is the diameter and l the length of the rod, while Δl is the elastic elongation caused by the tensile force. Plastic flow must not occur in this experiment and it is therefore necessary to keep $\Delta l/l$ very small; how small depends very much upon the material of the rod.

Since it is difficult to satisfy this condition and still attain great precision, in addition to the static method, Kundt (1865) already introduced a dynamic method of determining the modulus of elasticity. In the latter method the modulus of elasticity is calculated from the characteristic frequency of rods for longitudinal or for transverse vibrations [1]). In this case the relative changes in length may remain very small without detracting from the precision. The modulus of elasticity is connected with the characteristic frequency N_l of the longitudinal vibration of a cylindrical rod by the relation

$$E = \varrho\left(\frac{2}{n} N_l\, l\right)^2 \qquad (2)$$

In this expression ϱ is the density of the material and n a whole number which is 1 for the fundamental tone, 2 for the first overtone, etc.

From the transverse characteristic vibrations the modulus of elasticity can also be calculated. With a characteristic frequency N_{tr} of a square rod with the thickness a the following holds:

$$E = \varrho c\left(\frac{N_{tr}\, l^2}{a}\right)^2 \qquad (3)$$

c is here a constant which depends upon the order of the characteristic vibration and which has the value 0.947 for the fundamental tone.

Another elastic constant is the rigidity or modulus of shear Φ. We shall define this with the help of *fig. 1* in which $ABCD$ represents the

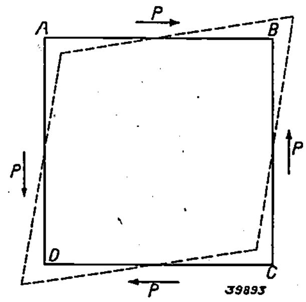


Fig. 1. Deformation of a cube by a shearing stress.

cross section of a cube of 1 cm³ of a given material. A force P acts along the upper surface, while along the other three surfaces BC, CD and DA forces also of the magnitude P act in the directions indicated. By these forces, which together form two opposite couples, the cube is deformed so that the cross section is now diamond-shaped with the angles $\pi/2-\varphi$ and $\pi/2+\varphi$. The rigidity Φ is then defined as P/φ.

The rigidity can also be determined by the measurement of the characteristic frequencies of elastic vibrations of a test rod. If N_t is the frequency of the torsional vibration (order n), of a circular rod, the following is true:

$$\Phi = \varrho\left(\frac{2}{n} N_t\, l\right)^2 \qquad (4)$$

The characteristic frequencies of the torsional vibrations in the case of a circular rod are thus independent of the diameter, as is the case for the longitudinal vibrations.

For the generation and measurement of the characteristic vibration electrical methods can now be used to advantage. We shall here describe an arrangement which was used in this labo-

[1]) In the static method the isothermal value of the elastic constants is found, in the dynamic method, the adiabatic value. The difference is slight (<1%) and can easily be calculated.

ratory and found suitable for the determination of the elastic constants of non-ferromagnetic metals. We shall then give several results which were obtained in measurements with this apparatus.

Method

We determined the elastic constants E and Φ of a number of non-ferromagnetic metals from the characteristic frequencies of the longitudinal and torsional vibration of the rods. We did not use the transverse vibration.

The rods were made to vibrate by placing them in the field of a permanent magnet and inducing alternating currents in the rods, whereupon the magnetic field exerts ponderomotive forces. *Fig. 2*

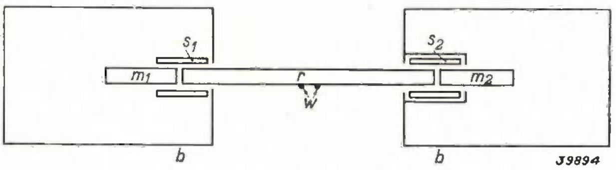


Fig. 2. Arrangement for the determination of the resonance frequency in the case of longitudinal vibrations. An alternating current of known frequency is sent through the coil s_1. Together with the inhomogeneous field of the magnet m_1 this current gives rise to longitudinal vibrations of the rod r. These vibrations, especially in the case of resonance, cause A.C. voltage in the coil s_2 which are observed with a cathode ray oscillograph.

is a diagram of the arrangement which was used for the excitation of the longitudinal vibration. The rod rests upon two thin metal wires w. Around the extremities of the rod are coils s_1 and s_2 and behind the rod the permanent magnets m_1 and m_2. Through coil s_1 flows an alternating current of variable and accurately known frequency which is generated by means of a tone generator. Eddy currents are induced in the rod r by these alternating currents. These eddy currents in the non-homogeneous field of the permanent magnet give rise to longitudinal forces which alternate with the frequency of the eddy currents. The frequency of the alternating currents is now varied until it is equal to one of the longitudinal characteristic frequencies of the rod, so that resonance occurs in this vibration form.

This mechanical resonance can be ascertained with the help of a cathode ray oscillograph. For that purpose a coil s_2 and a magnet m_2 are placed at the other end of the rod, in an arrangement identical to that of s_1 and m_1 for the excitation of the vibration. Because of the longitudinal motion of the second extremity of the rod in the magnetic field, eddy currents are also excited in this end of the rod which induce in coil s_2 an A.C. voltage which is fed, *via* an amplifier, to one of the deflection systems of the cathode ray oscillograph.

The coils are surrounded by earthed copper boxes b, while the rod is also earthed in order to make the mutual induction of the coils as small as possible. *Fig. 3* shows a photograph of the arrangement used.

Fig. 4 is a diagram of the arrangement used for torsional vibration, with the shielding boxes emitted for the sake of simplicity. The currents induced in the rod by the coils s_1, with the help of the field of the magnet m_1 give rise to a couple on the rod as a result of which torsional vibration may occur. In this case also the arrangement for the ascertainment of the vibration is similar to that for its excitation.

In order to make the occurrence of the couple clear in fig. 4 the eddy currents and the magnetic lines of force are indicated by arrows. The ponderomotive force which is perpendicular to the direction of the current and of the lines of force, acts, at the moment illustrated, forward above the

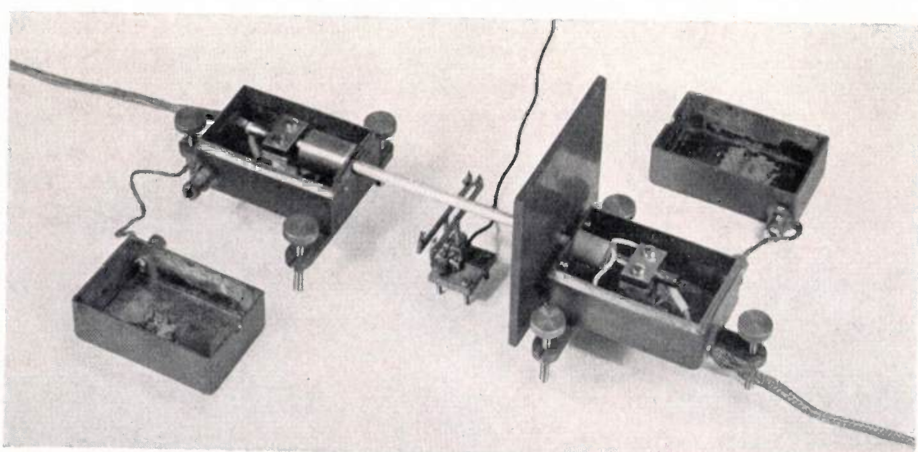


Fig. 3. Photograph of the arrangement represented in fig. 2.

axis of the rod and backward below the axis of the rod, so that a couple occurs in the direction of the arrow K. Due to the non-homogenity of the magnetic field the force directed toward the rear, which is excited close to the magnet, will be slightly larger than the force directed forward, so that in addition to the couple a tranverse force toward the rear is also exerted. In addition to the rotational vibrations therefore transverse vibrations may also be expected.

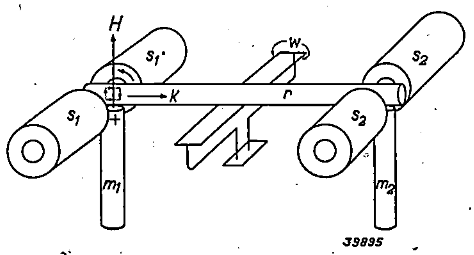


Fig. 4. Arrangement for the determination of the resonance frequency upon torsional vibrations; s_1, s_2 coils, m_1, m_2 magnets, w wires upon which the test rod r rests.

A possibility of ascertaining the vibrational form concerned consists in a comparison of the results, for rods of the same kind but of different diameter. The characteristic frequency of the transverse vibrations depends very closely upon the diameter, while that of torsional vibrations, as we have seen, is independent of the diameter. When the characteristic frequencies vary with the thickness of the rod, it may be concluded that one is not concerned with pure torsional vibrations.

Such a variation was actually found to occur when ferromagnetic rods were investigated. The fact that the transverse forces become especially large in this case is understandable in connection with the deformation of the magnetic field which the ferromagnetic rod causes. It was indeed found impossible to carry out reproducible elasticity measurements on these rods.

Several results

With the method described many characteristic frequencies could sometimes be measured. In the case of an aluminium rod for example, both for the longitudinal vibration and for the torsional vibration 8 overtones in addition to the fundamental tone could be measured. In the case of substances with less good electrical conduction and a greater internal damping than aluminium, sometimes only the fundamental could be measured.

In general the different overtones of a vibration furnish values of E as well as of Φ which agree very well among themselves and which are only slightly different for different test rods of the same metal. This is, however, not always the case. Thus in an investigation of the elastic constants in the series of alloys of copper and zinc [2]) remarkably divergent results were obtained in the case of one alloy, namely β-brass. The values found for E and Φ varied in a sometimes irreproducible way with the treatment which the rods had undergone. *Table I* gives an example of the results of several measurements. It was found that the chemical composition had only a slight effect on the elastic constants E and Φ and therefore that the differences which were found for the different rods, and which sometimes amount to 30 per cent or more, must be ascribed to other causes.

Table I

Modulus of elasticity E and rigidity Φ for β-brass (51-53% Cu, 49-47% Zn, sometimes 0.1% Cr) after different treatment.

Treatment	$E \cdot 10^{-11}$ dynes/cm²	$\Phi \cdot 10^{-11}$ dynes/cm²
not rolled	6·06	3·77
50% rolled at 700° C, annealed at 700°	6·54	3·43
30% rolled at 300° C, annealed at 500°	7·02	3·24
50% rolled at 500° C, annealed at 500°	9·46	3·52
8% rolled at 300° C, 30% rolled at 400° C, annealed at 450°	9·16	3·32

Explanation of the results

If the value of the modulus of elasticity is studied as it is measured by different authors for a given substance in polycrystalline state, it is striking that the differences which occur are considerably greater than could be explained by errors in measurement with the precision indicated, although they are seldom as great as in table I. These differences may be ascribed chiefly to poorly reproducible properties of the material used, such as:

1) impurities, deviations in the chemical composition,
2) cavities, cracks and similar material flaws,
3) anisotropy of the material.

As an example of 2) polycrystalline tungsten may be mentioned. The modulus of elasticity of drawn tungsten wires is found to depend upon their thickness, which may be ascribed to the fibrous structure of such wires. The cohesion in directions perpendicular to the wire is slight, so that in these directions the values found for the moduli of elasticity and the rigidity are too small.

Since the third point, anisotropy, may be very important for metals we shall go into it in somewhat

[2]) M. J. Druyvesteyn and J. L. Meyering, Physica, 8, 1059, 1941.

more detail. A polycrystalline material is built up of a number of crystals. For most single crystals the modulus of elasticity depends very closely upon the direction in which it is measured. A polycrystalline material will only have a modulus of elasticity which is independent of the direction (*i.e.* will be elastically isotropic) when the crystals are distributed uniformly over all directions.

In the case of cast test rods this condition will sometimes be satisfied. It may, however, also happen that upon solidification of the liquid material a certain preferential direction of the crystallites becomes evident (so-called casting texture [3]). Upon working a metal, by hammering, forging or drawing, new textures may appear. In all these cases the elastic constants depend upon the state of working and on the direction in which the constants are measured; the scattering in the values of E and Φ which hereby occur may amount to more than 30 per cent.

The occurrence of a texture can also be ascertained by taking X-ray photographs of the substance. It is, however, also possible to ascertain the presence of preferential orientations from the elasticity measurements themselves. For this purpose, in addition to the moduli of elasticity and the rigidity, the compressibility K is also determined. K indicates the relative change in volume divided by the pressure from all sides necessary to produce it. Furthermore the Poisson constant μ must be determined (*i.e.* the relation between the relative lateral contraction to the relative elongation which occurs in stretching a rod).

For an isotropic homogeneous substance the following is valid:

$$K = \frac{9}{E} - \frac{3}{\Phi}, \qquad \mu = \frac{E}{2\Phi} - 1. \quad . \quad . \quad (5)$$

If E, Φ, K and μ are measured the correctness of these formulae can be checked and in this way it can be ascertained whether or not the material investigated was isotropic and homogeneous. Since a slight change in E or Φ usually leads to a large change in the calculated values of μ and K, this is a very sensitive method.

[3]) Moreover in the case of cast material there is a relatively large chance of the occurrence of cavities.

[4]) E. Grüneisen, Ann. Physik 22, 801, 1907.
D. A. G. Bruggeman, dissertation, Utrecht 1930.

β-brass

Since the elastic anisotropy of β-brass single crystals is abnormally great, a texture will in this case lead to great differences in the elastic constants E and Φ, so that the explanation of the differences occurring in table I becomes obvious. In *table II* in addition to the values of E and Φ from table I we also give the values calculated from those values according to equation (5) for K and μ. By direct measurements Bridgman found for the compressibility of β-brass a value of 9.1×10^{-13} dyne^{-1} cm^2, while the value of μ for every isotropic substance must lie between zero and 0.5, and usually lies between 0.15 and 0.45. On the basis of the values calculated for K and μ it may therefore be concluded that the test rods used for the first three substances were certainly not homogeneous and isotropic.

Table II

Modulus of elasticity E and rigidity Φ, together with the values calculated according to equation (5) of the compressibility K and the Poisson constant μ for β-brass.

$E \cdot 10^{-11}$	$\Phi \cdot 10^{-11}$	$K \cdot 10^{13}$	μ
Dynes/cm^2	Dynes/cm^2	Dynes^{-1} cm^2	
6.06	3.77	70	—0.19
6.54	3.43	51	—0.05
7.02	3.24	35	0.09
9.46	3.52	9.9	0.34
9.16	3.32	7.5	0.38

In the last two measurements the calculated value of K does not deviate very much from the value measured by Bridgman and the calculated value of μ is not contradictory to the theoretically expected value. The elastic constants of the polycrystalline isotropic material must therefore be deduced from these measurements, so that it may be stated that the modulus of elasticity of β-brass probably lies between 9.0 and 9.5×10^{-11} dynes/cm^2.

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

An adequate number of reprints for the purpose of distribution is not available of those publications marked with an asterisk. Reprints of other publications may be obtained on application to the Natuurkundig Laboratorium, N.V. Philips' Gloeilampenfabrieken, Eindhoven (Holland), Kastanjelaan.

1554: E. J. W. Verwey: Energieën in de colloidchemie (Energies in colloid chemistry) (Handelingen Ned. natuur- en geneesk. Congr., Apr. 1941, p. 88-90).

A survey is given in this lecture of the attempts to calculate the hydration and solvation energies of electrolytic solutions for positive and negative ions separately on the basis of the theory of Bernal and Fowler which has been somewhat improved by the author. These quantities are very important for boundary surface phenomena in colloid chemistry. The investigations will be published in detail in Rec. trav. chim. 1941.

1555: A. Bouwers: De biologische and medische toepassingen van kunstmatige radioactiviteit (The biological and medical applications of artificial radioactivity) (Handelingen Ned. natuur- en geneesk. Congr., Apr. 1941, p. 207-211).

In this lecture before a combined meeting of physicists, biologists and medical men a survey was given of the method by which artificial radio active substances can now be prepared by means of transmutation of atomic nuclei (*cf.* Philips techn. Rev., 6, 46, Feb. 1941). These substances have already been used with much success in biological and medical research. In order to measure the intensity of the radioactivity of a substance or in certain parts of the organism to be examined, the biologist or doctor can profitably use a simple electron counter which was demonstrated during the lecture and which is also described in detail in Philips techn. Rev. 6, 75, Mar. 1941. In conclusion the favourable results were pointed out which have also already been obtained with artificial radioactivity in the field of therapy.

1556: L. Blok: Toestel voor het electrisch meten van de gehoorscherpte (Apparatus for the electrical measurement of the acuity of hearing) (Handelingen Ned. natuur- en geneesk. Congr., Apr. 1941, p. 212-214).

For the contents of this lecture before a combined meeting of physicists and doctors the reader may be referred to an article which has since appeared in Phil. techn. Rev. 6, 234, Aug. 1941 on this audiometer.

1557: B. D. H. Tellegen: Phaenomenologie der piezo-electriciteit (The phenomena of piezo-electricity) (Ned. T. Natuurk. 8, 270-274, July 1941).

In this opening address of the symposium on piezoelectricity organized by the Netherlands Physical Society and the Netherlands Radio Society (May 1941) a description is first given, on the basis of a one-dimensional image, of how the electrical polarization and the mechanical deformation of crystals may be connected. The equations are then derived thermodynamically which describe these piezoelectric phenomena, while finally it is indicated how these considerations must be extended in order to treat a three-dimensional case, such as actually occurs.

1558: J. de Boer: Toepassing van piezo-electrische kristallen bij geluidsweergave. (The application of piezoelectric crystals in sound reproduction). Ned. T. Natuurk. 8, 345-356, July 1941).

In this address at the symposium on piezoelectricity (May 1941) the way is discussed in which mechanical vibrations can be converted into electrical A.C. voltages and the reverse with the help of crystals of Rochelle salt. The peculiar piezoelectrical properties of such crystals, such as the occurrence of two Curie temperatures, are discussed. Finally forms of construction for microphones and gramophone pick-ups with piezoelectric crystals are discussed (*cf.* also: Philips techn. Rev. 5, 140, 1940).