VOL. 12 No. 1, pp. 1-32 JULY 1950

Philips Technical Review

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

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

A SMALL GREENHOUSE WITH ARTIFICIAL LIGHTING FOR STUDYING PLANT GROWTH UNDER REPRODUCIBLE CONDITIONS

by R. van der VEEN.

631.544:581.14.001.5

Experiments carried out under partly unknown conditions have in many cases led to results which are valueless or difficult to interpret, the trouble and time expended on them thereby being lost. The same is the case in the field of plant-physiology, where so many factors are of influence. With the greenhouse described here all the main factors are well under control, so that it will undoubtedly prove valuable for obtaining reproducible results in this field.

The development of a plant depends upon a a number of factors, the most important of which are light (its intensity and colour, and also the length of time during which the plant receives light), temperature, the amount of moisture and of carbon dioxide in the air, and the composition of the soil. In order to study the effect of these factors upon plant growth it must be possible to control each one of them separately. Where plants are allowed to grow naturally there can be hardly any question of such a control, particularly in a capricious climate, so that the results of any investigation undertaken will scarcely be reproducible. With the hothouses or greenhouses commonly used it is possible to exercise some control over conditions but not by any means to the extent desired; think, for instance, of the sunlight, which is so unreliable, and the temperature, often rising in the summer higher than one would wish.

Now, in order to make it possible to study the behaviour of plants under well controllable conditions, Philips Laboratories have built a number of "greenhouses" in which the entirely artificial lighting can be adjusted in intensity, spectral composition and duration, and the temperature, the humidity and the CO_2 content of the air can be controlled within certain limits. *Fig. 1* is a sketch of such a greenhouse having a floor space of 1 m × 1.5 m and a height of 1.5 m. Here a description will be given of these greenhouses, one of which is depicted in *fig. 2*.

Light

The lighting is by means of tubular fluorescent lamps (TL lamps) of 40 W, of which there are eleven on each of the two sides and eight at the top. This arrangement gives a very favourable distribution of the light: small plants, or the bottom leaves of larger ones, are not overshadowed by tall plants or, respectively, by the top leaves. When all thirty lamps are alight the intensity of the light is approximately equal everywhere inside the hothouse, amounting to about 60-70 W/m² (10,000-12,000 lux).

The lamps can be switched on and off in groups of three, so that one can work with lower intensities. The lamps on the two sides alone yield more than sufficient light for the normal growth of most plants, the wattage then amounting to about 1 kW per 1.5 m² floor area. Only in the case of some tall-growing plants, like the Poinsettia, was it found desirable to use also the top lamps.

As to the spectral composition of the light, this can be greatly varied by employing different kinds of TL lamps. In the first place there are the normal types available, known as "warm-tone", "white" and "daylight", the spectrum of the first-mentioned type containing a relatively large amount

of red and yellow, whilst that of the "daylight" type contains a fair amount of blue; the colour temperatures of these three types are respectively about 2900, 4000 and 6500 °K. Then there are other, special, TL lamps that can be used, which are lined with phosphors producing, for example, a bluish, greenish or reddish light. By mounting different lamps in the same greenhouse a light can be obtained in practically any desired tint.

Temperature

Although TL lamps radiate comparatively little heat, the temperature inside the greenhouse might rise to too high a level (up to 35 °C with normal ambient temperature) if there were no means of carrying off the heat. Provision has therefore been made for this by arranging for tap-water to flow down the outer side of the glass panes separating the TL lamps from the space within which the plants

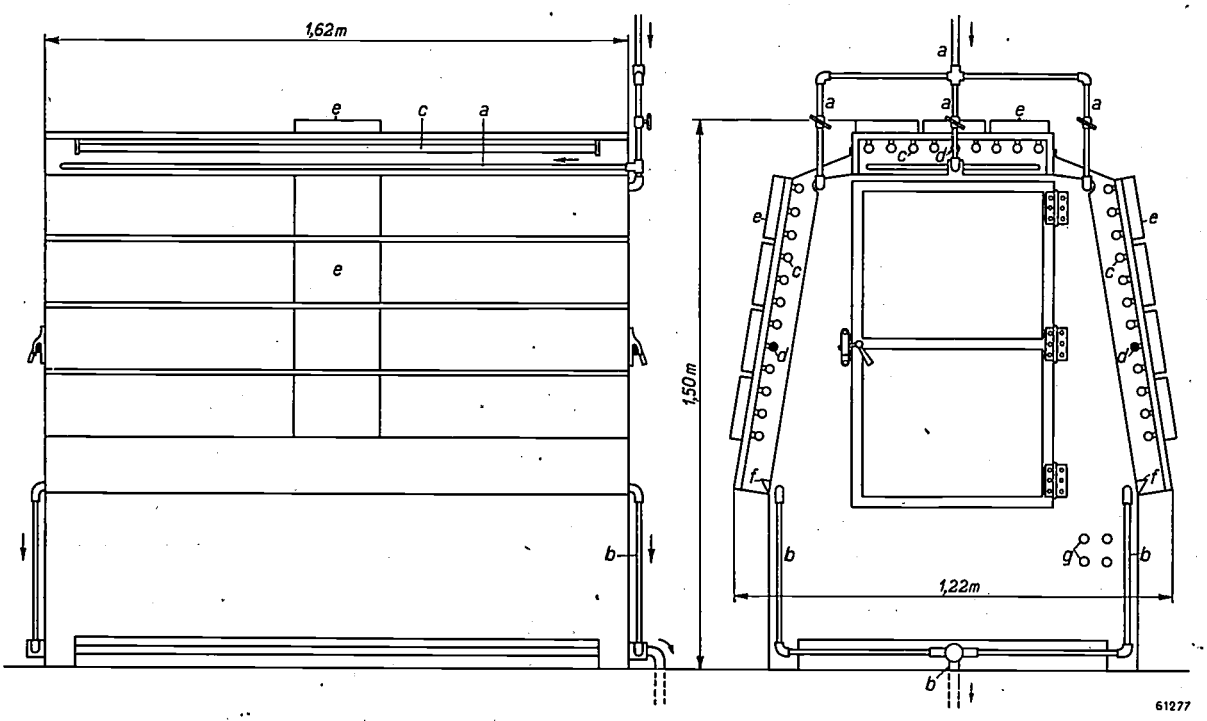


Fig. 1. Sketch of the greenhouse, front and side views. *a* inlet of pipe water for cooling, *b* water outlet, *c* TL lamps (30 in number), *d* mercury lamps (three), *e* starters for the lamps (in groups of three), *f* discharge channel for the cooling water, *g* openings for carrying off condensation water and inlets for air, for CO_2 and for the electric current for heating.

In the designing of the greenhouse account has been taken not only of the possibility of lighting it with TL lamps, but also of the possibility of lamps being used which radiate too much infra-red for plants, such as high-power incandescent lamps. For the latter case provision has been made for a relatively deep layer of water to be introduced at the top of the greenhouse with a view to absorbing the greater part of the infra-red rays coming from the lamps installed above the water.

Another very important factor is the length of time during which the plants are daily exposed to the light [1]). This period of time can be set to any length by means of a time-switch (seen in fig. 2).

[1]) For some of the problems relating to this photoperiodicity see, for instance, R. van der Veen, Influence of light upon plants, Philips Techn. Rev. 11, pp. 43-49, 1949/50 (No. 2).

are growing. By means of cocks in the inlet pipes the temperature inside the greenhouse can be regulated between 35 °C and 17 °C. While the TL lamps are switched off the temperature level can be maintained by switching on some incandescent lamps mounted under the floor of the greenhouse so that their light does not reach inside.

At first some trouble was experienced from algae growth in the cooling water. These algae come from germs which are always present in tap-water. They form a green deposit on the glass and make it necessary to clean the panes thoroughly at least once a week. This algae growth has now been prevented by mounting between the TL lamps on each of the two sides and at the top a mercury lamp (TUV lamp, *d* in figs 1 and 3), the ultra-violet rays from which kill the algae; the plants inside the greenhouse are not affected by these ultra-violet

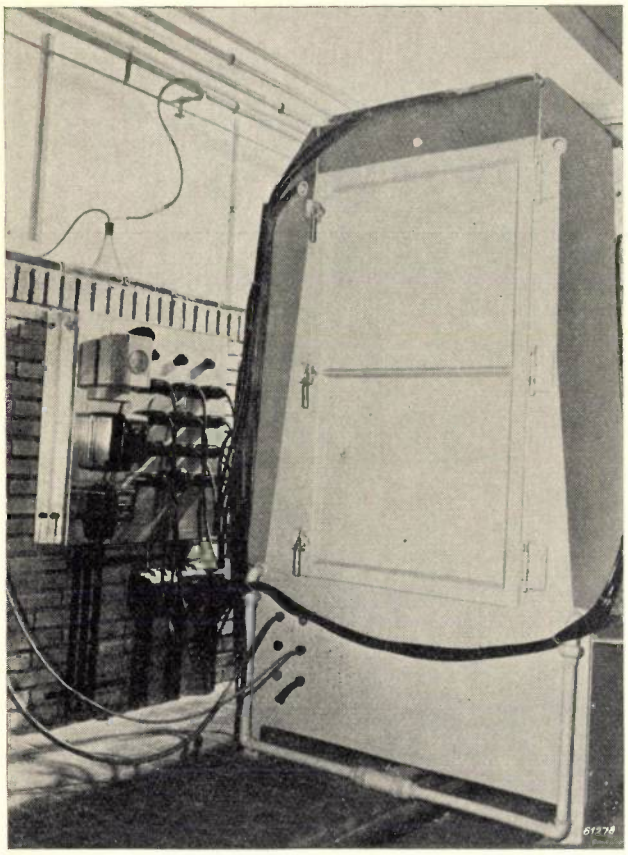


Fig. 2. Front of the greenhouse. On the left the switchboard with plug sockets to which the TL lamps can be connected in groups of three, and with a time-switch for regulating the daily exposure to light. On the left at the bottom: supply lines for air and carbon dioxide.

rays because the glass panes hold them back.

Owing to the cooling there is a great deal of condensation on the inside of the glass panes. The water from this condensation is collected in gutters leading outside the greenhouse; it appears to be free of algae germs, so that — once the panes have been thoroughly cleaned — there is not usually any algae growth, notwithstanding the absence of ultra-violet rays on the inside.

Humidity and CO_2 content of the air; nature of the soil

The greenhouses are connected to a compressed-air line supplying dry air. The more of this air that is allowed to flow through the greenhouse, the lower is the degree of humidity, so that in this way the moisture content of the air inside can be controlled. The normal flow of air supplies an amount of carbon dioxide that is usually sufficient for the growth of the plants, but the CO_2 concentration can be increased when required from a carbon dioxide line likewise connected to the greenhouse. Both lines can be fed from steel cylinders.

Of course attention has also to be paid to the soil in which the plants are set. This factor, however, can be eliminated by applying techniques of water culture, for which provision has also been made in the designing of these greenhouses. One is then independent not only of the climate but also of the soil.

Some results

It is not the place here to go deeply into the results obtained, but some examples may be given to illustrate what can be attained with these greenhouses.

The first experiments were made with cucumber, tomato, Chrysanthemum, Poinsettia, Phlox, tobacco, lettuce and balsam plants. All these kinds of plants showed normal development and were cultivated in the greenhouses from seedlings to flowering and fruit-bearing plants (see, for instance, the

Fig. 3. A greenhouse with Coleus and Petunias. A TUV lamp (*d*) is to be seen between the TL lamps. The ultra-violet rays from this mercury lamp kill the algae which otherwise form a deposit on the glass panes along which the cooling water flows.

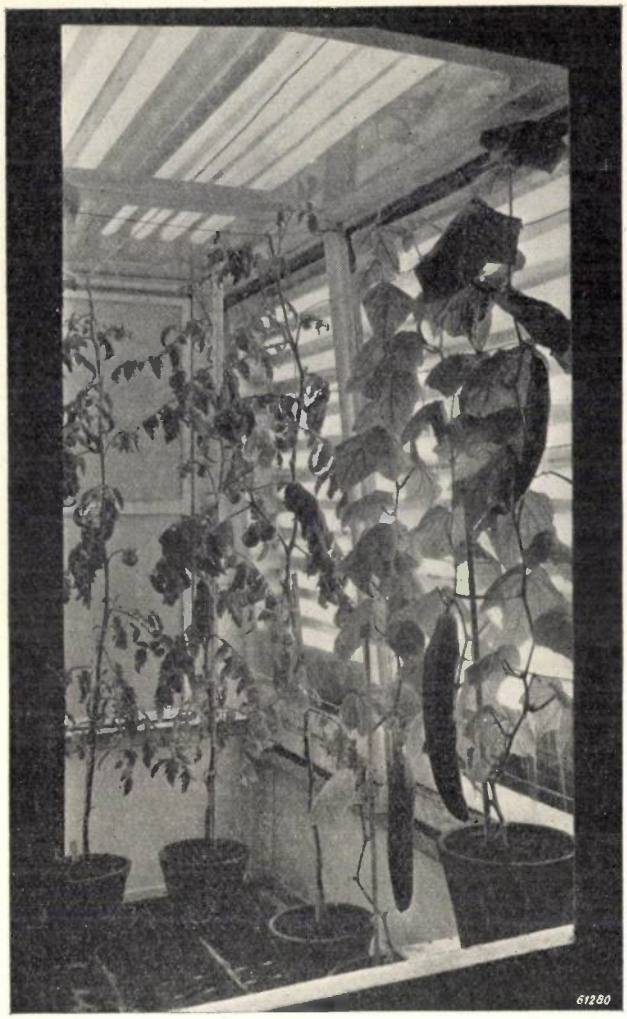


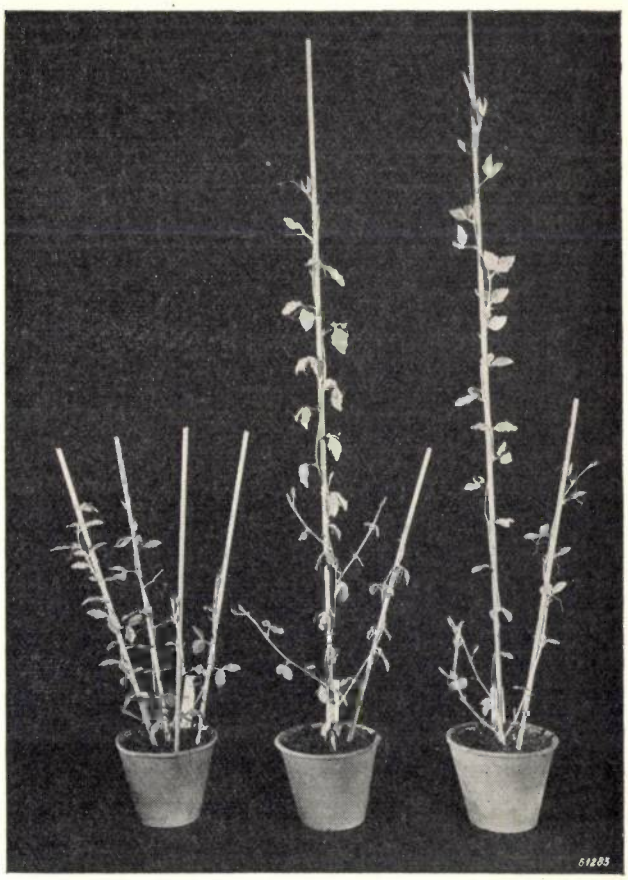
Fig. 4. Flowering and fruit-bearing cucumber and tomato plants cultivated from seedlings in the artificial climate of the greenhouse.

tomato and cucumber plants illustrated in *fig. 4*). The influence of the colour of the light is clearly demonstrated by the tomato and sweet-pea plants in *fig. 5*: lack of blue rays results in tall but spindly plants with few branches; light that is rich in blue rays, on the other hand, produces stocky, strong specimens with an abundance of offshoots.

Finally *fig. 6* gives an example of the influence of photoperiodicity upon Coleus, a typically long-day plant (see the article quoted in footnote [1]). Daily exposure to long hours of light (or continuous light as applied in the case of the specimen on the left in the photograph) promotes the development of this plant and gives the leaves a beautiful dark-red colour with a bright-green edge. Nine hours of light per day resulted in a less sturdy plant and the colours were paler.

For botanical and horticultural laboratories a number of greenhouses as described here will prove a useful and valuable addition to the equipment available, especially where physiological experiments are conducted. These greenhouses also lend themselves excellently for experiments in connection with photoperiodicity (see the article referred to in footnote [1]).

a



b

Fig. 5. *a*) Tomato plants exposed to light of a different spectral composition, but otherwise cultivated under the same conditions and of the same age. The plant on the left has been exposed to a predominantly blue light, the middle one to green and the one on the right to red light (the last two with a deficiency of blue). Blue light checks the growth but produces sturdy plants with plenty of offshoots and dark-green leaves. A deficiency of blue light makes the plants grow spindly, with light-green leaves. *b*) The same for sweet peas.

Of course it goes almost without saying that these greenhouses are not intended for horticulture. For nursery work it is generally more advantageous to use normal glasshouses, where full benefit can be derived from sunlight and in which artificial lighting is only employed in the winter months to make up for the shortage of daylight.

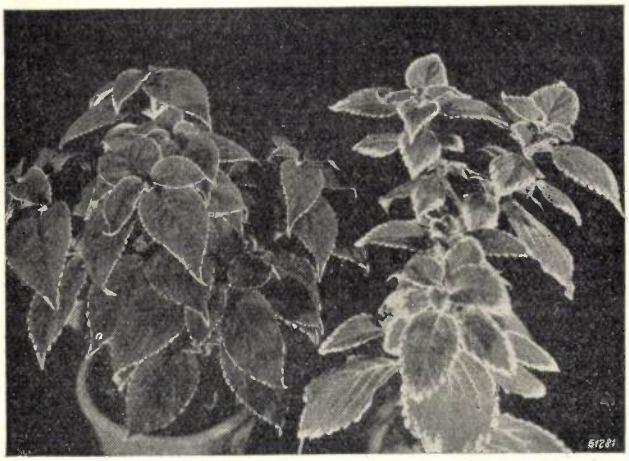


Fig. 6. Two equally old Coleus plants (a species of long-day plant), the one on the left exposed to continuous light and that on the right to nine hours light per day but otherwise treated in the same way. The specimen continually exposed to light is stockier and has better coloured leaves.

Summary. In order that experiments can be carried out in regard to the processes of plant growth under well-defined conditions, a greenhouse has been constructed in which a number of factors affecting the plant are fully under control. The entirely artificial lighting is derived from 30 TL lamps giving a uniformly distributed light of 60 to 70 W/m² (10,000 to 12,000 lux) in a space of 1 m × 1.5 m × 1.5 m. Practically any spectral composition can be obtained by using TL lamps lined with different phosphors. A time-switch provides for an adjustable daily exposure to light. The temperature inside the greenhouse can be controlled between 17 °C and 35 °C by regulating a supply of water from the mains which is made to flow down the outside of the glass walls. Algae growth in the cooling water is prevented by the ultra-violet radiation from three mercury lamps (TUV lamps). When the lighting is switched off the temperature can be maintained by electrical heating. The humidity and the CO_2 content of the air can be separately regulated. Experiments with a large number of species have shown that under these artificial conditions seedlings can be raised successfully to flowering and fruit-bearing plants.

THE INFLUENCE OF TEMPERATURE ON THE FLUORESCENCE OF SOLIDS

by F. A. KRÖGER and W. de GROOT.

535.371

The efficiency of the fluorescence of solids has become of particular importance in illumination engineering since fluorescent lamps came into use. In this connection it is of special interest to know how this property varies with temperature. Generally speaking it appears that efficiency decreases with rising temperature, i.e. at high temperatures fluorescence disappears. A study of this quenching is of importance for gaining an insight into the process of fluorescence, especially when the efficiency is measured in combination with the delay time as a function of temperature.

Introduction

The fluorescence of gases, solutions and solids under irradiation with visible or ultra-violet rays [1] [2] [3] is closely related to the absorption of that light in the system irradiated. The absorption of a quantum of light (photon) in an atom or molecule of a gas, in an ion or molecule in solution or in a crystal increases the energy of the system, lifting an electron into a higher quantum state. What happens further with such an "excited" electron depends upon the circumstances. In the simplest cases it returns to its original state, the energy obtained from the absorption of a photon being wholly or partly radiated again in the form of a photon. This phenomenon is called fluorescence, and in the case where the absorbed energy is entirely radiated again one also speaks of resonance.

It often happens, however, that the "excitation energy" obtained from absorption is not converted into a photon but into some other form of energy. This may even be the case with a mono-atomic gas. An example is the quenching of the resonance radiation in sodium vapour by the addition of a foreign gas, such as nitrogen. When an excited Na-atom collides with an N_2-molecule there is a chance of the excitation energy being transmitted to that molecule and transformed into kinetic energy, or into vibrational or rotational energy.

In the case of multi-atomic gases and of molecules in solution the process of fluorescence is likewise often replaced by a radiation-less process, the excitation energy being converted into vibrational energy, dissociation energy, or energy required for a chemical reaction, etc.

In the case of acetone vapour it is known, for instance, that only three fluorescence photons are emitted against 100 absorbed ultra-violet photons, the excitation energy from the remaining 97 photons being converted in one way or another into molecular vibrations.

In the case of solutions such a "quenching" may arise when an excited molecule collides with a non-excited molecule of the same kind or with a molecule of the solvent. Therefore, in order to get a highly fluorescent solution a solvent has to be chosen whose molecules have little or no influence upon the excited molecules of the fluorescent substance. There may also be chemical reactions competing with the process of fluorescence, a known example of this being the extinction of the fluorescence of an aqueous solution of fluorescein by KI, whereby an $I(H_2O)^-$ ion is split into $I + H + OH^-$ at the cost of the excitation energy of a fluorescein molecule.

Fluorescence and quenching in the case of solids

In the case of solids the situation is rather more complex. In some cases a fluorescent solid contains atoms or ions, as for instance those of the rare earths or manganese, or groups of atoms, like UO_2 or WO_4, which absorb light selectively. This includes both simple compounds, e.g. CeF_3, $UO_2(NO_3)_2$, $CaWO_4$, and mixed crystals such as $(Zn,Mn)_2SiO_4$. To these belong also substances like ZnS, which become fluorescent by the addition of a small quantity of a foreign metal (Cu, Ag) usually incorporated in the sulphide lattice as chloride (CuCl, AgCl). In the case of sulphide phosphors the permissible concentration of the admixture is very small (10^{-5} to 10^{-1} atom % per molecule of the base material), whereas with manganese phosphors this concentration is greater (0.1 to 10

[1] W. de Groot, Philips Techn. Rev. **3**, 125-132, 1938.
[2] J. H. Gisolf and W. de Groot, Philips Techn. Rev. **3**, 241-247, 1938.
[3] F. A. Kröger, Philips Techn. Rev. **6**, 349-358, 1941.

at. %). The active atom in the admixture (such as Mn or Cu in the examples given) is called the activator.

When such a substance is irradiated and the rays are absorbed at a wavelength within the selective absorption range of the atoms, ions or groups of atoms referred to, then the process of fluorescence resembles very much that taking place in a gas or in a solution. Quenching of the fluorescence might occur here when interaction takes place between the excited atom or ion or the activated group of atoms and the surrounding crystal lattice. The energy accumulated in the centre of fluorescence owing to the absorption of a photon is then entirely converted into vibrational energy of the neighbouring atoms of the crystal lattice and ultimately into vibration of the whole lattice, thus into heat.

Every crystalline substance, whether or not it contains atoms or groups of atoms capable of acting as centres of fluorescence, shows absorption in a certain wavelength range due to the crystal lattice as such and not to any particular centres. When a photon is absorbed in such a wavelength range then an electron of the crystal lattice is brought into a "band" of greater energy. In a few cases also this energy may immediately be radiated again in the form of a photon (an example is the green fluorescence of CdS at —180 °C). In the case of the phosphors just mentioned, however, the excitation energy absorbed by the crystal lattice is mostly first transmitted from the crystal lattice to one of the activator centres present in the lattice.

Sometimes, in addition to the activator, there is also a second kind of atom, the sensitizer, with its own specific absorption. The energy absorbed by this atom is transmitted to the activator centre. An example is calcium phosphate with Ce (sensitizer) and Mn (activator), or calcium silicate with Pb(s) and Mn(a). The sensitizer may itself also cause fluorescence. In the cases mentioned for instance both Pb and Ce fluoresce in the ultra-violet.

With solids, therefore, there may be a great variety of mechanisms of fluorescence due to one or more non-radiating transmission processes taking place between the primary process (absorption of a photon) and the ultimate process of fluorescence (the emission of a photon). Moreover there may also be non-radiating processes whereby the whole of the excitation energy is converted into lattice energy, resulting in quenching and absence of fluorescence. The counteraction between the useful and the useless non-radiating processes determines the quantum efficiency (the number of fluorescent photons per photon absorbed). Since, as will be shown, the probability of any non-radiating process taking place is a certain function of the temperature, the quantum efficiency may either increase or decrease with rising temperature. Given a sufficiently high temperature, quenching processes as a rule predominate, so that ultimately the efficiency is reduced, and at very high temperatures even drops to zero.

Closely related to the influence of temperature upon the quantum efficiency is a second phenomenon, namely the influence of temperature upon the variation of the intensity of the fluorescence with time after irradiation has ceased, thus upon the decay.

What has been dealt with above qualitatively will now be explained quantitatively with the aid of some simple theoretical hypotheses, whilst something will also be said about the measurement of the quantities and the conclusions to be derived from the results.

The relation between efficiency and decay[4])

Suppose we have a centre irradiated with the wavelength that is absorbed by that centre. We then have the situation as represented in *fig. 1*. Through absorption of a quantum an electron is raised from the original state (*0*) to a higher quantum state (*1*). The energy of the centre increases from E_0 to E_1. Let it be possible for the centre

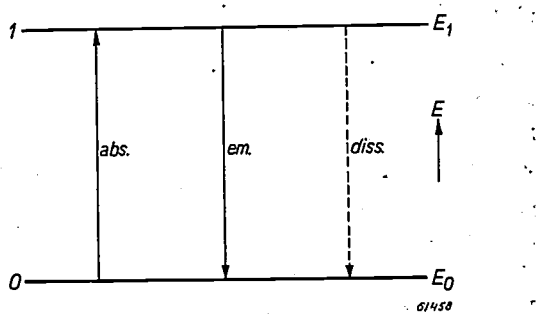


Fig. 1. Absorption and emission in the case of a single excited state (*1*). In addition to the emission (———) also a radiation-less transition (— — —) is indicated. E = energy.

to return to the state *0* under the emission of a photon. If there is no other possibility then the quantum efficiency equals 1, i.e. a photon is emitted for every photon absorbed.

[4]) It is expressly pointed out that in this article by "decay" is meant only the, usually short, decay of the fluorescence and not the, mostly long, decay of phosphorescence, which is governed by other factors and, for instance, is dependent upon the temperature in an entirely different way.

In fig. 1 it has been assumed, for the sake of simplicity, that the energy of the photon emitted is equal to that of the absorbed photon. This is the resonance, already mentioned, occurring in the case of mono-atomic gases (e.g. sodium vapour).

In all other cases, as with molecules and especially with solids, the situation is usually more complicated (see *fig. 2*). Owing to the absorption of a photon the centre undergoes a change into the state *1*. Usually this absorption is immediately

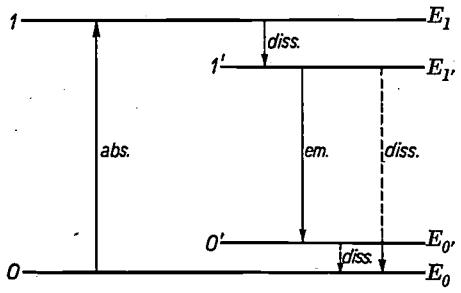


Fig. 2. Explanation of the difference in frequency between emission and absorption in composite systems (molecules in solution, and solids; Stokes's rule). Radiation-less processes are represented by dotted lines.

followed by a non-radiating transition to a lower state of energy *1'*. The states *1* and *1'* may, for instance, be two vibrational states belonging to the same electron state. The emission process will be a transition from the state *1'* to the state *0'*, which usually differs from *0*. This is immediately followed by the transition $0' \rightarrow 0$, again without emission.

The energy differences $E_1 - E_{1'}$ and $E_0 - E_{0'}$ are imparted to the crystal lattice as vibrational energy. As a result the emitted quantum $h\nu_{em}$ is less than the absorbed quantum $h\nu_{abs}$ ($\nu_{em} < \nu_{abs}$ or $\lambda_{em} > \lambda_{abs}$, Stokes's rule). This results in a decrease of the energy efficiency in the ratio ν_{em}: ν_{abs}, even if each absorption process leads to the emission of a photon. The quantum efficiency however remains equal to 1.

Let us now consider the process of fluorescence as a function of time. The number of centres in the state *1* is denoted by n and the number of photons absorbed per second by A. Further, for an electron in state *1* the probability of a transition to the original state under the emission of a photon in the interval of time dt is taken to be equal to $\gamma_f dt$. For n we then have the equation

$$dn = A\,dt - \gamma_f n\,dt,$$

or

$$\frac{dn}{dt} = A - \gamma_f n. \quad \ldots\ldots \quad (1)$$

When irradiation is continued a long time, a stationary state is reached for which $dn/dt = 0$ and thus n is constant ($= n_0$):

$$n_0 = A/\gamma_f.$$

The number of photons emitted per unit of time is then $\gamma_f n_0 = A$, as is obvious, since the quantum efficiency is 1.

When the irradiation is stopped at a certain moment and the time is counted from that moment onwards ($t = 0$) then for $t > 0$

$$\frac{dn}{dt} = -\gamma_f n. \quad \ldots\ldots\ldots \quad (2)$$

and

$$n = n_0 \exp(-\gamma_f t). \quad \ldots\ldots \quad (3)$$

The number of photons emitted per unit of time then amounts to

$$\gamma_f n = \gamma_f n_0 \exp(-\gamma_f t) = A \exp(-\gamma_f t). \quad (4)$$

Thus the fluorescence decreases exponentially with time, and the greater the value of γ_f, the more rapid is the decay.

This case shows a strong resemblance to the acoustical problem of reverberation [5]) and, further, to a large number of physical phenomena based on a formula similar to (2). In particular we would mention, in view of what follows, the discharge of a capacitor with capacitance C and charge Q via a resistance R, where

$$\frac{dQ}{dt} = -\frac{Q}{CR}$$

and thus $Q = Q_0 \exp(-\delta t)$, with $\delta = 1/RC$. A suitable measure for the duration of the decay or, in the case of the capacitor, the duration of the discharge, is the reciprocal value of the constant occurring in the exponential function (thus $1/\gamma_f$ and $1/\delta$ respectively), i.e. the time in which the quantity in question has diminished to $1/e = 0.37$ of its original value.

Let us assume that in addition to the fluorescence transition a dissipation process is also possible, with probability constant γ_d. This means that the probability of an excited centrum giving off its energy to the crystal lattice, without emission, in the interval of time dt, is $\gamma_d dt$. Then, with constant irradiation,

$$\frac{dn}{dt} = A - (\gamma_f + \gamma_d)\,n. \quad \ldots\ldots \quad (5)$$

In the stationary case ($dn/dt = 0$) the number of excited centres is now $n_0' = A/(\gamma_f + \gamma_d)$ and the number of photons emitted per unit of time is

$$\gamma_f n_0' = A\,\frac{\gamma_f}{\gamma_f + \gamma_d}.$$

The quantum efficiency now apparently is

$$\eta = \frac{\gamma_f}{\gamma_f + \gamma_d} = \frac{1}{1 + \gamma_d/\gamma_f} < 1.$$

[5]) See W. Tak, Philips Techn. Rev. 8, 82-88, 1946.

At the same time there is also a change in the time law governing the decay of the fluorescence after the irradiation has stopped, for then

$$\frac{dn}{dt} = -(\gamma_f + \gamma_d)\, n,$$

so that

$$n = n_0' \exp\left[-(\gamma_f + \gamma_d)\, t\right].$$

The number of photons emitted per unit of time is

$$\gamma_f n = \gamma_f n_0' \exp\left[-(\gamma_f + \gamma_d)\, t\right] = A\eta \exp(-\varkappa t), \quad (6)$$

in which $\varkappa = \gamma_f + \gamma_d$.

The duration of the decay is now $1/\varkappa$, thus less than it was before. Once the quantities η and $\varkappa$ that have to be determined experimentally have been measured, it is easy to calculate therefrom the quantities γ_f and γ_d that are of importance for the theory. One finds:

$$\gamma_f = \varkappa\eta, \quad \ldots\ldots\ldots \quad (7)$$

$$\gamma_d = \varkappa(1-\eta). \quad \ldots\ldots\ldots \quad (8)$$

The arguments set forth above were applied a long some time ago by Stern and Volmer to the case of the fluorescence of gases and liquids. Their application to the case of the fluorescence of solids is of relatively recent date [6]).

of a certain wavelength, is projected by a second quartz lens L_2 onto the specimen P. This specimen consists of a thin layer of the substance under test applied to a metal plate which can be either heated electrically or given a low temperature by means of a cooling bath. The specimen is contained in a vacuum bulb of quartz with a view to keeping the temperature as constant as possible and in order to prevent condensation of water vapour on the specimen at low temperatures. The fluorescent light is picked up by a concave mirror Sp and reflected upon a photocell C via a filter F that does not transmit the original ultra-violet radiation. The incident light is measured by replacing P by a plate of magnesium oxide, of course without using the filter F.

With this arrangement it is possible to measure in relative units the intensity of the fluorescent light as a function of the temperature and that of the incident ultra-violet radiation. The ratio of these intensities gives a quantity that is proportional to the efficiency, the "relative efficiency". It is much more difficult to determine the efficiency in absolute units, but in most cases this can be dispensed with, the maximum value of the quantum efficiency being simply taken as equal to 1. For the information required the absolute value of

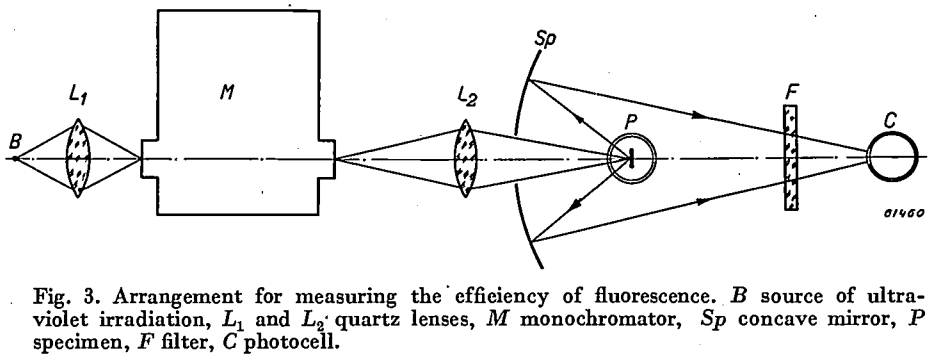


Fig. 3. Arrangement for measuring the efficiency of fluorescence. B source of ultra-violet irradiation, L_1 and L_2 quartz lenses, M monochromator, Sp concave mirror, P specimen, F filter, C photocell.

Measurement of the efficiency

Fluorescence efficiency is measured with the set-up represented diagrammatically in *fig. 3*. The image of a source B of ultra-violet rays, for instance a high-pressure mercury lamp, is projected by a quartz lens L_1 onto the input aperture of a quartz monochromator M. The image of the output aperture of M, which thus emits ultra-violet rays

the efficiency plays a less important part than the temperature-dependency of the relative efficiency.

The difficulty in determining the absolute efficiency arises from various circumstances. For instance, it is not certain whether the angular distribution of the fluorescent light will be the same as that of the light diffusely reflected by the magnesium-oxide plate. Further, for an absolute measurement the photocell has to be replaced by a bolometer, the sensitivity of which is not dependent upon the wavelength. But even when all this is taken into account there is still an element of uncertainty, because part of the energy absorbed is not used for the process to be studied but is lost in some other way. This will certainly be the case, for instance, when the specimen is mechanically mixed with a neutral absorbing powder (say soot black), but it may also occur with a pure specimen.

[6]) L. Brüninghaus, Ann. Physik (6) 2, 55-75, 1948. N. A. Tolstoj and P. P. Feofilov, Doklady Ak. Naoek 59, 235, 1948. K. H. Hellwege, Naturwissenschaften 31, 212-213, 1947. See also the articles quoted in footnote [10]).

On the other hand part of the fluorescent light will be lost in the specimen through absorption. In many cases, however, if disturbing influences are properly taken into account it is found that within a considerable temperature range, in particular at low temperatures, the measured quantum efficiency of the processes studied very closely approximates to unity.

Measurement of the decay

The decay of the fluorescence as a function of time after interruption of the irradiation is measured in the set-up depicted in *fig. 4*, which in many respects resembles that of fig. 3 but is provided with rotating discs having openings cut in them, so that the whole arrangement shows a marked resemblance to the known Becquerel phosphoroscope. The image of the source B is now projected thus at a moment

$$t = \frac{\varphi}{2\pi} \cdot t_s ,$$

where t_s represents the time of the revolution of the discs. Each time the slit passes through the beam a flash of light is allowed to pass through and this can be measured separately, for instance with the aid of a cathode-ray oscillograph, or integrated over the time (thus as a direct current) by means of a galvanometer shunted by a capacitance. By varying the angle φ (for this purpose S_2 has been made adjustable with respect to the shaft) the intensity of the afterglow can be studied as a function of t. In many cases a time function is found as in equation (6), thus an intensity proportional

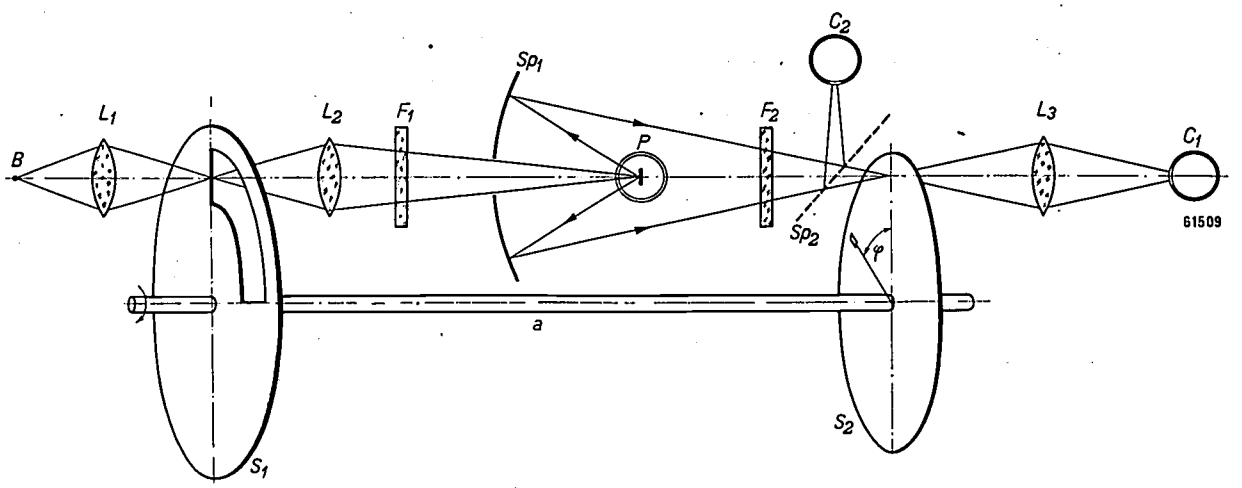


Fig. 4. Arrangement for measuring the decay of fluorescence. B source of ultra-violet irradiation, L_1 and L_2 quartz lenses, L_3 glass or quartz lens, P specimen, C_1 and C_2 photocells, S_1 and S_2 discs rotating on a common shaft a, Sp_1 concave mirror, Sp_2 flat mirror, F_1 and F_2 filters.

onto the specimen P by two lenses, L_1 and L_2. Here the monochromator is replaced by a filter F_1. The lens L_1 casts a picture of B onto the disc S_1, which has an opening of 90° and thus lets the rays through during one quarter of a revolution and shuts them off during three quarters of that time. In the diagram the instant is depicted when the light has just been shut off and thus the decay begins. The image of the fluorescent strip of the specimen P is formed on the photocell C_1 via the mirror SP_1 and the lens L_3. At the place where the mirror forms an intermediate image is a second disc, S_2, mounted on the same shaft as S_1. This second disc has a narrow radial slit which can be so adjusted that the light is only passed through at the moment that S_1 has revolved through an angle φ from the beginning of the interruption, to exp $(-\varkappa t)$. The logarithm of the intensities found is then plotted in the known way as a function of t, and $\varkappa$ is determined from the slope of the straight line drawn through the points of measurement.

Another way to find the intensity as a function of t is to arrange for the light coming from the concave mirror to fall, via a second, flat mirror Sp_2, upon the photocell C_2, the signal from which is fed to the pair of vertical deflection plates of a cathode-ray oscillograph. When the other pair of deflection plates is connected to a linear time base then the screen shows both how the fluorescence increases at the beginning of the irradiation and how it decays after irradiation has ceased, it then being possible to measure and study both these time functions on the screen. When the fluorescence decays according to an exponential function of

time then by this method the time constant $\varkappa$ can be determined directly, by applying an exponential time base. This is done by connecting the horizontal deflection plates of the oscillograph to a capacitor (capacitance C) that is periodically charged and discharged via a resistance R at the rate of the revolution of S_1. The horizontal deflection is then proportional to exp $(-\delta t)$, where $\delta = 1/RC$. If $\delta = \varkappa$, which can be attained by adjusting R, then a straight line is produced on the screen, whilst in all other cases the line is curved. In this way $\varkappa$ can be found very quickly [7]).

A certain, simple case will be taken to show what results are obtained from the combined measurements of efficiency and intensity. As an example we have chosen the fluorescence of ammonium uranyl pentafluoride, $(NH_4)_3UO_2F_5$, irradiated with ultra-violet rays of the wavelength $\lambda = 3650$ Å. The fluorescence of this compound is due to the UO_2 group. Both the absorption and the emission spectra consist of a group of narrow bands [8]), the emission bands having wavelengths between 4700 and 6050 Å. In *fig. 5* the logarithm of the intensity of the fluorescence after the irradiation has ceased has been plotted as a function of t for a number of temperatures between —148 °C and +123 °C. It is seen that a number of straight lines are obtained, which proves that the decay

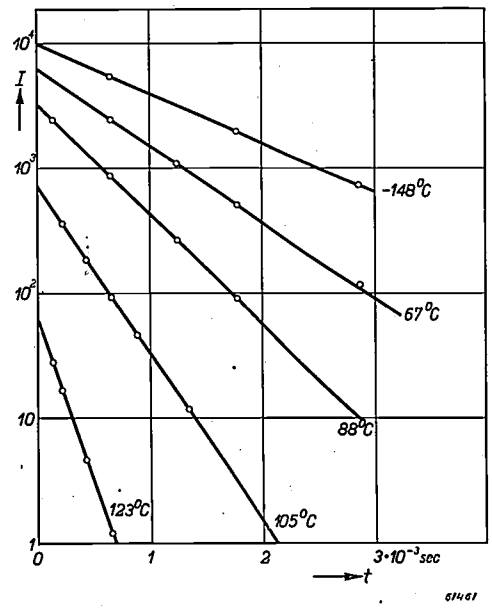


Fig. 5. Intensity of the fluorescent light plotted on a logarithmic scale as a function of time after interruption of the irradiation.

[7]) This method of measuring resembles very much a method already described in this journal for determining the acoustical reverberation time; see the article quoted in footnote [5]).

[8]) See also Philips Techn. Rev. 3 (1938), p. 131.

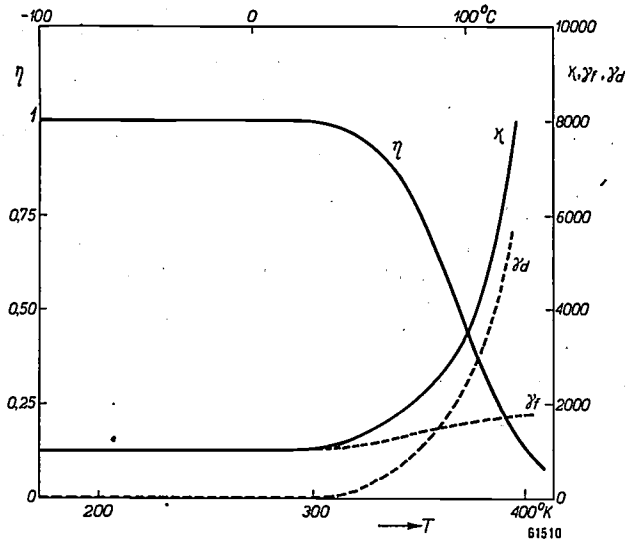


Fig. 6. The efficiency η and the reciprocal decay time $\varkappa$ for ammonium uranyl pentafluoride as a function of the temperature T. The quantities γ_f and γ_d are also plotted.

follows the exponential law $I_t = I_0 \exp(-\varkappa t)$ and also that $\varkappa$ increases with rising temperature. In *fig. 6* $\varkappa$ has been plotted as a function of the absolute temperature T, together with the efficiency of the fluorescence η. Here, for the reasons given above, the efficiency at low temperatures has been taken as being equal to unity. From the values of $\varkappa$ and η the values of γ_f and γ_d have been calculated according to the formulae (7) and (8), and these quantities have likewise been indicated in fig. 6.

Discussion of the measurements

It is seen that γ_f varies but little with T and can, therefore, to a first approximation be regarded as a constant. On the other hand the quantity γ_d, which is practically negligible at temperatures below 20 °C, increases rapidly with T at higher temperatures. In order to investigate what law this increase follows, log γ_d has been plotted in *fig. 7* as a function of $1/T$. It appears that, within the errors of observation, the variation of log γ_d can be represented by the straight line

$$\log \gamma_d = 12.48 - \frac{3570}{T}.$$

Thus γ_d can be represented by an exponential function of the shape

$$\gamma_d = S \exp(-\varepsilon/kT), \quad \ldots\ldots \quad (9)$$

where k is Boltzmann's constant ($k = 1.38 \times 10^{-23}$ joule/°K, $1/k = 11.600$ °K/eV) and $S = 3 \times 10^{12}$, whilst $\varepsilon = 0.7$ eV.

This points to the fact that a centre already in an excited state requires a certain additional

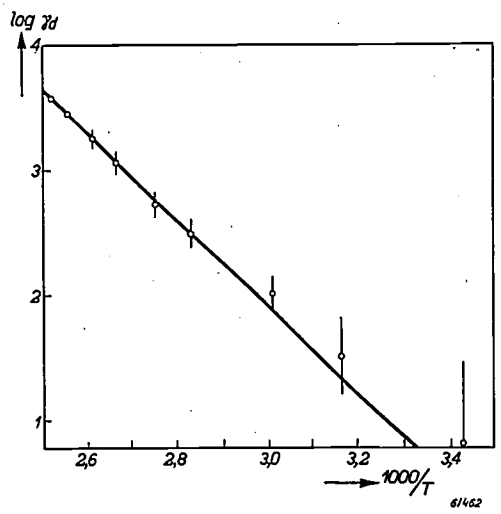


Fig. 7. Logarithm of the dissipation constant γ_d as function of $1/T$ (T = absolute temperature). The vertical lines denote the inaccuracy of the measurements.

energy ε (activating energy) for the non-radiating transition from excitation energy into lattice energy to take place. This hypothesis was introduced long ago by Jablonsky and has more recently been applied by Mott and Gurney to the case of solids. The highly complicated problem was greatly simplified by substituting one imaginary coordinate x for the large number of coordinates corresponding to the numerous degrees of freedom of the centre in the surrounding crystal lattice.

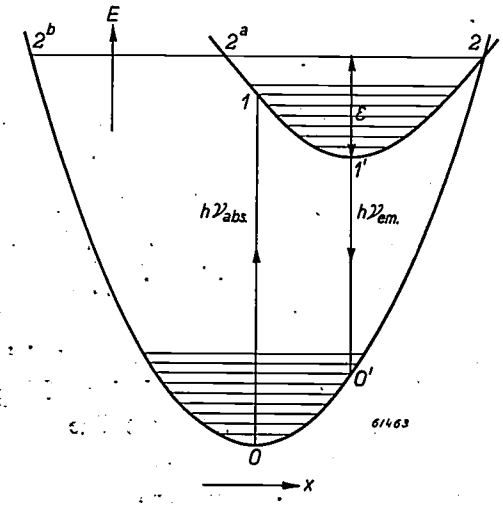


Fig. 8. Model depicting the theory of the quenching of fluorescence in the case of solids, according to Mott and Gurney.

In *fig. 8* the potential energy of the system: centre + immediate surroundings (briefly referred to in the following as "centre") has been plotted as a function of x both for the original state and for the excited state. The minimum of the potential curve represents in each case a state of equilibrium around which vibrations are possible. A number of vibrationary states have been indicated both for the original state and for the excited state. Further there are indicated the energies $h\nu_{abs}$ and $h\nu_{em}$ of an absorbed and an emitted photon (compare fig. 8 with fig. 2). After the excitation $0 \rightarrow 1$ the centre quickly passes into the state $1'$, whereby the vibration energy is imparted to the surrounding crystal lattice. Inversely, in the state $1'$ the centre may also draw energy from the crystal lattice, its vibrational energy then increasing. Upon the point 2 being reached, which shows an energy difference ε with respect to $1'$, it is then possible for the centre to pass from the state in which it oscillates between 2 and 2^a into a state which belongs to the lowermost potential curve and represents a very high vibrational condition of the ground state (oscillation between 2 and 2^b). As a consequence of the coupling between the vibrations and the rest of the lattice, which does not find expression in this model, this vibrational energy will then very soon spread over the entire crystal lattice, so that the state 0 is reached without a photon having been emitted. Since the probability for the transition $1' \rightarrow 2$ is proportional to $\exp(-\varepsilon/kT)$, formula (9) is theoretically justified and the physical significance of the "activation energy" ε is explained.

The behaviour described here for the case of the uranyl salt has also been found in a large number of other cases, two of which may be mentioned here, whereby a manganese ion acts as centre of fluorescence, viz: Mg_2TiO_4 activated with Mn present as a quadrivalent ion (Mn^{4+}), and the well-known willemite (Zn_2SiO_4-Mn_2SiO_4), in which Mn occurs as a bivalent ion (Mn^{2+}).

We shall now briefly deal with some cases of fluorescence of solids where the behaviour is more complicated than can be described with the aid of the theoretical hypothesis given above.

Complicated mechanisms of fluorescence

To gain some insight into more complicated cases of fluorescence we must first follow some theoretical considerations. Suppose that (*fig. 9*) in addition to the original state (0) there are at the same time two states of excitation (1) and (2) corresponding to the energies E_1 and E_2 ($E_2 > E_1$), and further that owing to the absorption of a photon there is an electron in the state 2. We now assume that transitions are possible from 2 to 0, from 2 to 1 and from 1 to 0. Each transition may take place either with the emission of a photon or without any radiation. The probability of these processes in an interval of time dt is represented as:

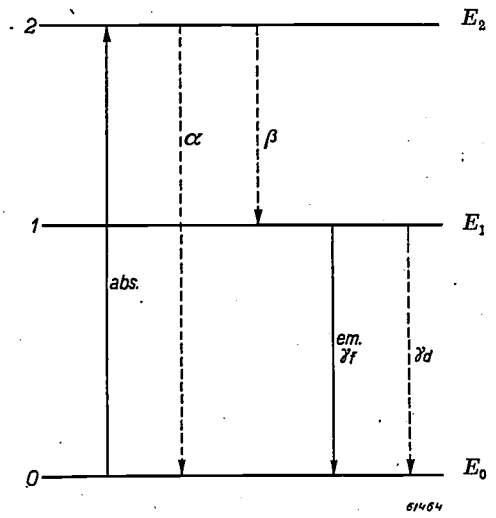


Fig. 9. System with two excited states (*1* and *2*), indicating the transition through absorption to state *2*, the transition $1 \rightarrow 0$ under emission, and various radiation-less transitions (in dotted lines).

	with radiation	without radiation
$2 \rightarrow 0$	$\alpha_f\, dt$	$\alpha_d\, dt$
$2 \rightarrow 1$	$\beta_f\, dt$	$\beta_d\, dt$
$1 \rightarrow 0$	$\gamma_f\, dt$	$\gamma_d\, dt$

In many cases no radiations are found corresponding to α_f and β_f, or in other words α_f and β_f are zero, or at least $\alpha_d \gg \alpha_f$, $\beta_d \gg \beta_f$. In the simplified assumption that this is the case, we find for the efficiency of the fluorescence (transition $1 \rightarrow 0$):

$$\eta = \frac{\beta_d}{\alpha_d + \beta_d} \cdot \frac{\gamma_f}{\gamma_f + \gamma_d} \quad \ldots \quad (10)$$

If, as is often the case, γ_f is small compared with α_d and β_d, it may easily be deduced that α_d and β_d have no effect upon the decay of the fluorescence after interruption of the irradiation but that an increase of γ_d accelerates that decay. If with such systems there is found to be a decrease in efficiency then this opens the possibility of investigating at what level the interfering radiation-less process takes place: if it is found that with increasing quenching (decreasing value of η) there is at the same time a decrease in the duration of the decay then the quenching is caused by γ_d; if that is not the case then the cause lies in the ratio of α_d/β_d. Examples of both cases have been found experimentally. Systems are also known where both effects take place simultaneously.

From formula (10) it appears that the efficiency is determined not only by the ratio γ_d/γ_f but also by that of α_d to β_d. If α_d and β_d increase differently with the temperature it is quite possible that at first, before the influence of γ_d is noticeable, the efficiency increases with rising temperature. This possibility has already been pointed out in the introduction.

The foregoing also applies in the case where state *2* does not belong to the centre as such but represents an excited state of the crystal lattice. This means that then an electron travels through the lattice, just as is the case with the positive charge (electron "hole") that is left after the electron is removed from a lattice ion (see, e.g., the article quoted in footnote [2]). For the transition from state *2* to state *1* it is now necessary that the electron and the positive charge meet near a centre, a part $E_1 - E_0$ of the energy $E_2 - E_0$ being thereby transmitted to the centre, while the remainder $(E_2 - E_1)$ goes to the lattice. The number of such transitions per unit of time is proportional to the product of the concentrations of excited electrons and the holes, thus proportional to the square of the number n_2 of electrons in the state *2*, contrary to the cases considered above, where the number of transitions was invariably proportional to the first power of the number of electrons in the respective state. The most important consequence of this is that, if such a "bimolecular" process governs the variation of the fluorescence as function of time during the decay, this variation is no longer exponential but "hyperbolic", which means to say that for high values of t the intensity decreases proportionally [9]) with $1/t^2$.

If in this case, in addition to the transition leading to fluorescence, there is a quenching process, likewise governed by a "bimolecular" process (recombination of electrons and holes), then we find again for the efficiency an expression of the form:

$$\eta = \frac{\beta}{\alpha + \beta} \cdot \frac{\gamma_f}{\gamma_f + \gamma_d}.$$

Here β is the coefficient determining the transition from the state *2* to state *1* and α a similar coefficient for the extinction process $2 \rightarrow 0$.

Still more complicated is the case where the radiation-less transition from *2* to *1* is indeed "bimolecular", i.e. proportional to n_2^2, but the competing process, the radiation-less transition $2 \rightarrow 0$, is "monomolecular", i.e. proportional to the first power of n_2. Obviously the counteraction between these processes then depends upon the value of n_2. In such a case the efficiency will be found to be dependent upon the intensity of the irradiation.

Such a behaviour has indeed been found with willemite (Zn_2SiO_4—Mn_2SiO_4) containing a relatively-large amount of manganese: a preparation containing,

[9]) Cf. Philips Techn. Rev. 3, 241-247, 1938 (especially pp 246 and 247).

say, 5% manganese can be made to fluoresce by irradiating it with a wavelength of 3650 Å, this wavelength being selectively absorbed by the Mn-ions. The decay of the fluorescence after the irradiation has ceased is exponential with time, the decrease in efficiency with rising temperature beginning at the same point where also the decay begins to decrease. Just as has been explained for the case of the uranyl salt, here too γ_f and γ_d can be determined ($S = 8.7 \times 10^8$, $\varepsilon = 0.89$ eV). When, however, the system is irradiated with $\gamma = 2537$ Å, coming within the range of lattice absorption, the behaviour of the decay remains the same but the efficiency begins to decrease at lower temperatures, the more so as the intensity of the irradiation is less (*fig. 10*). Something similar is found in the case of willemite containing, in addition to say 1% Mn_2SiO_4, a very small quantity (0.1%) of Fe_2SiO_4, the Fe-ions acting as quenchers.

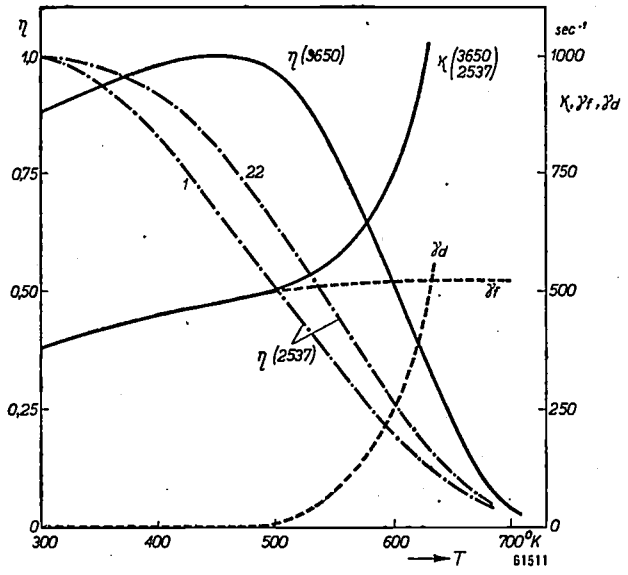


Fig. 10. The efficiency η and the quantities $\varkappa$, γ_f and γ_d for willemite with 5% Mn as function of the temperature T under irradiation with $\gamma = 3650$ Å and $\gamma = 2537$ Å, in the latter case for two intensities in the ratio 1 : 22.

These brief indications of such more complicated cases have to suffice here. For a more detailed treatment of the subject reference is made to the original publications dealing with the investigations [10]).

[10]) F. A. Kröger, W. Hoogenstraaten, M. Bottema and Th. P. J. Botden, The influence of temperature quenching on the decay of fluorescence, Physica **14**, 80-96, 1948.
F. A. Kröger and W. Hoogenstraaten, Decay and quenching of fluorescence in willemite, Physica **14**, 425-441, 1948.
W. Hoogenstraaten and F. A. Kröger, The intensity-dependence of the efficiency of fluorescence of willemite phosphors, Physica **25**, 541-556, 1949 (No. 5/6).
For older literature, dealing especially with the fluorescence of gases and liquids, see P. Pringsheim, Fluorescence and Phosphorescence, Interscience P.C., New York, 1949.

Summary. It is explained how the efficiency of the fluorescence of all kinds of substances is influenced by dissipative processes tending towards quenching of the fluorescence. First a simple case is taken to show the relation that must exist between the increasing quenching with rising temperature and the simultaneous decrease of the decay time. Methods are briefly discussed for measuring the efficiency and the duration of the decay, and the results of the measurements for simple cases are compared with the theory. Finally mention is made of the complications arising in less simple cases and examples are given of cases where the efficiency depends not only upon the temperature but also upon the intensity of the irradiation.

A MEASURING ARRANGEMENT FOR WAVEGUIDES

by A. E. PANNENBORG.

621.392.26: 621.317.3

In installations for generating and receiving microwaves the super- and ultra-high-frequency energy is carried by waveguides. The construction of such waveguide systems and the accessory apparatus has to meet demands quite different from those required of transmitting and receiving installations for longer waves. Special measuring arrangements are needed to ascertain whether these demands are satisfied.

Introduction

The spectrum of electromagnetic oscillations can be divided into two parts. To the first part, which we shall call the optical range, belong all oscillations having a wavelength less than 0.01 mm, whilst in the second part, the electrical range, lie the oscillations with wavelengths greater than approximately 1 mm. Oscillations with intermediate wavelengths play no part so far in practical applications. (The limits mentioned here cannot be sharply defined and are to be regarded only as indicating the order of size.) The short-wave end of the electrical range, the so-called microwave range, with wavelengths between 1 mm and about 30 cm, can in a certain sense be regarded as a transitional range, and it is this part of the spectrum with which we shall deal in this article.

The last decade has seen great progress in the development of microwave technique; its application in radar and other related apparatus has become generally known. Since the wavelength of microwaves is comparable to the dimensions of normal circuit elements (capacitors, coils, etc.) such as are commonly used for longer waves, these elements can no longer be employed in a radar installation and we have to use elements of an entirely different nature. Neither is it possible for the microwave energy to be transported via a normal parallel-wire transmission line, because the losses due to radiation would then be prohibitive, so that either coaxial cables or waveguides have to be used. The theory of the propagation of electromagnetic waves through such systems is well known. It appears that a coaxial cable can transport microwave energy in a suitable manner only when half the sum of the circumferences of the inner and outer conductors is less than the wavelength of the oscillation. For wavelengths exceeding 10 cm the coaxial cable is highly suitable, but for smaller wavelengths when, as in radar installations, the power to be transported is, moreover, considerable (sometimes even amounting to several megawatts peak) it is advantageous to use waveguides. These have a much greater power-carrying capacity than the small coaxial cables required for centimetric waves, without sparking occurring in the guide, and moreover in this range they have reasonable cross-sectional dimensions, of the order of the wavelength of the oscillation.

When waveguides are used as transmission lines all sorts of problems are encountered. To mention some of them: there is the manner in which the energy has to be fed into the waveguide at one end and taken off at the other end, the effect of bends and junctions, the way in which the electromagnetic field is influenced by obstacles in the guide, the matching of the various parts of a waveguide system, and so on. In order to illustrate the great importance of these problems in practice, a diagram (*fig. 1*) is given, representing the high-frequen-

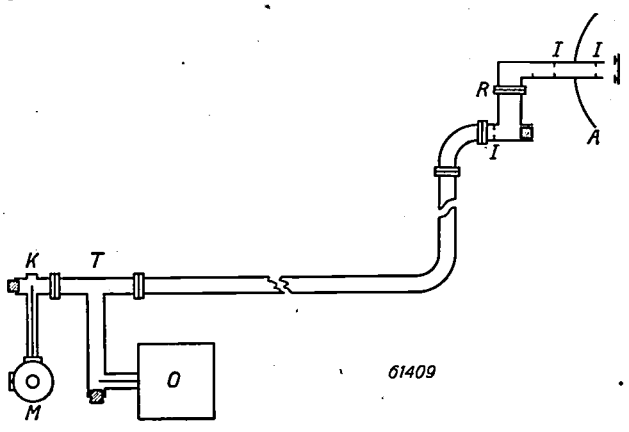


Fig. 1. Lay-out of the high-frequency part of a radar installation. The oscillations generated by the magnetron M are conducted by a coaxial cable and a coupling K to the waveguide, which carries them to the aerial A, where they are radiated into space. This aerial serves at the same time for receiving the reflected signals, for instance from an aeroplane, which pass along the same waveguide to the receiver O, connected to the waveguide via a branch T. The aerial is rotated with the aid of a rotating joint R. To allow for correct matching some diaphragms I occupy part of the cross section of the waveguide. (From Principles of Radar, M.I.T. Radar School Staff, McCraw Hill Book Cy, New York 1946.)

cy part of a radar installation, clearly showing the bends, couplings, etc. occurring in it.

For investigating the properties of these obstacles, bends, etc. it is necessary to take certain measurements. Measurements of the reflections are particularly important, because, as we shall see presently, in a waveguide reflections generally have to be avoided. It is these measurements that will be dealt with in this article; naturally only a limited part of the vast measuring technique can be discussed and for a more intensive study the reader is referred to the extensive literature already published on this subject [1]).

Before proceeding to describe the technique of measuring we have to give a brief review of the essential theory of waveguides. This theory has already been dealt with more extensively in two articles previously published in this journal [2]).

Propagation of waves in waveguides

Here we shall confine ourselves to rectangular waveguides with sides a and b, where $b < a$. As described in article I, in such a waveguide modes of different types may occur. There are transverse electric modes of various orders (TE_{mn} modes) and transverse magnetic modes of various orders (TM_{mn} modes). It appears that a TE_{mn} mode (we are dealing here only with TE modes) can be propagated through the guide only when the following relation between the wavelength λ and the dimensions of the guide is satisfied:

$$\frac{1}{\lambda} > \frac{1}{\lambda_{cr}} = \sqrt{\left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2}; \quad \ldots \quad (1)$$

λ_{cr} is called the critical wavelength for the TE_{mn} mode.

From formula (1) it immediately follows that the greatest critical wavelength is given by $\lambda_{cr} = 2a$, belonging to a TE_{10} mode, whilst the next greatest critical wavelengths are given by $\lambda_{cr} = 2b$ (TE_{01} mode) and $\lambda_{cr} = a$ (TE_{20} mode).

It is desirable that the waves should be propagated in a guide in a single mode, since the reflecting properties of a discontinuity differ for each mode of oscillation. The guide is therefore given such dimensions that the mode of oscillation lies between the greatest and the next greatest critical wavelength, i.e. so that only TE_{10} waves can be propagated in the guide. The bandwidth available is made as large as possible by choosing $b < {}^1/_2\, a$; it is then determined by the length of the wide side a; all microwaves having a wavelength satisfying the inequality $a < \lambda < 2a$ can then be propagated through the guide as a TE_{10} mode.

Waves with a greater wavelength cannot penetrate into the guide, because as soon as they enter it they change into a stationary wave (perpendicular to the axis) the amplitude of which diminishes exponentially along the axis of the guide. There is also some attenuation in the case of the propagation of waves within the permissible frequency band, owing to the dissipative losses caused by the current flowing along the surface of the guide; the closer the wavelength lies to the greatest critical wavelength, the greater is the attenuation. In practice, therefore, only the wavelength range $a < \lambda < 1.6a$ can be used. See also article I, pp 24 and 25.

The formulae describing the electromagnetic field of a TE_{10} mode in a rectangular waveguide are (see I, p. 23 and 24):

$$E_y = A \sin\frac{\pi x}{a} e^{2\pi j\varphi},$$

$$H_x = A \sqrt{\frac{\varepsilon_0}{\mu_0}} \sqrt{1 - \left(\frac{\lambda}{2a}\right)^2} \sin\frac{\pi x}{a} e^{2\pi j\varphi},$$

$$H_z = A \sqrt{\frac{\varepsilon_0}{\mu_0}} \frac{\lambda}{2a} \cos\frac{\pi x}{a} e^{2\pi j(\varphi - 1/4)},$$

$$E_x = E_z = H_y = 0,$$

where

$$\varphi = \frac{c}{\lambda}\left\{t - \frac{z}{c}\sqrt{1 - \left(\frac{\lambda}{2a}\right)^2}\right\}.$$

In these formulae ε_0 and μ_0 represent respectively the dielectric constant and the magnetic permea-

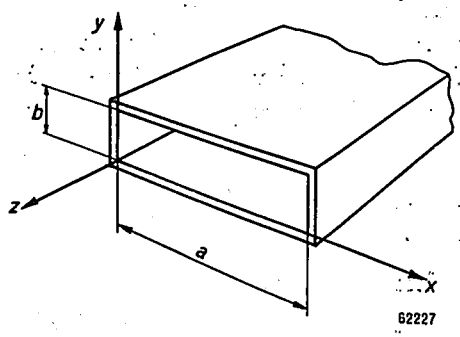


Fig. 2. Sketch of a rectangular waveguide with sides a and b ($a > b$). The system of coordinates indicated in the sketch is that used for the calculations in this article.

[1]) See, e.g., C. G. Montgomery, Technique of microwave measurements, Radiation Laboratory Series 11, McGraw Hill Book Cy, New York, 1947, and J. G. H. Huxley, A survey of the principles and practice of wave guides, Cambridge University Press, 1947.

[2]) W. Opechowski, Electromagnetic waves in wave guides, I, Philips Techn. Rev. 10, 13-25, 1948, and II, Philips Techn. Rev. 10, 46-54, 1948. From now on these articles will be referred to as I and II respectively.

bility of the vacuum, λ represents the wavelength in vacuo, c the velocity of light and A an arbitrary constant.

The system of coordinates has been so chosen that the z-axis is parallel to the longitudinal direction of the wave guide, the x-axis parallel to the side a and the y-axis parallel to the side b of the cross section (*fig. 2*).

From the above formulae it is seen that the z-dependency of the mode is given by the factor $\exp(-2\pi jz\sqrt{1/\lambda^2-(1/2a)^2})$. This means that the wavelength λ_z at which the oscillation is propagated in the waveguide is not equal to λ but is given by the formula

$$\frac{1}{\lambda_z}=\sqrt{\frac{1}{\lambda^2}-\frac{1}{(2a)^2}} \quad \ldots\ldots \quad (2)$$

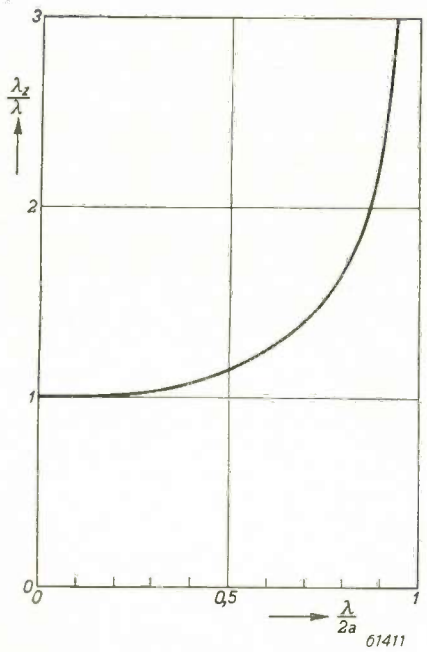


Fig. 3. The wavelength λ_z of an oscillation propagated in a waveguide of the width a, divided by the wavelength λ of the oscillations as produced by the generator, plotted as a function of the ratio $\lambda/2a$. It appears that λ_z may easily be twice as great as λ.

In *fig. 3* λ_z/λ has been plotted as a function of $\lambda/2a$. It is seen that λ_z can easily be twice as great as λ.

Fig. 4a shows the magnetic lines of force for the case of a TE_{10} wave in a rectangular waveguide; here the significance of λ_z is made quite clear. In this diagram the electric lines of force are at right angles to the plane of drawing, the intensity of the electrical field varying across the width of the waveguide as indicated in fig. 4*b*. Finally the system of currents flowing along the surface of the waveguide is represented in fig. 4*c*.

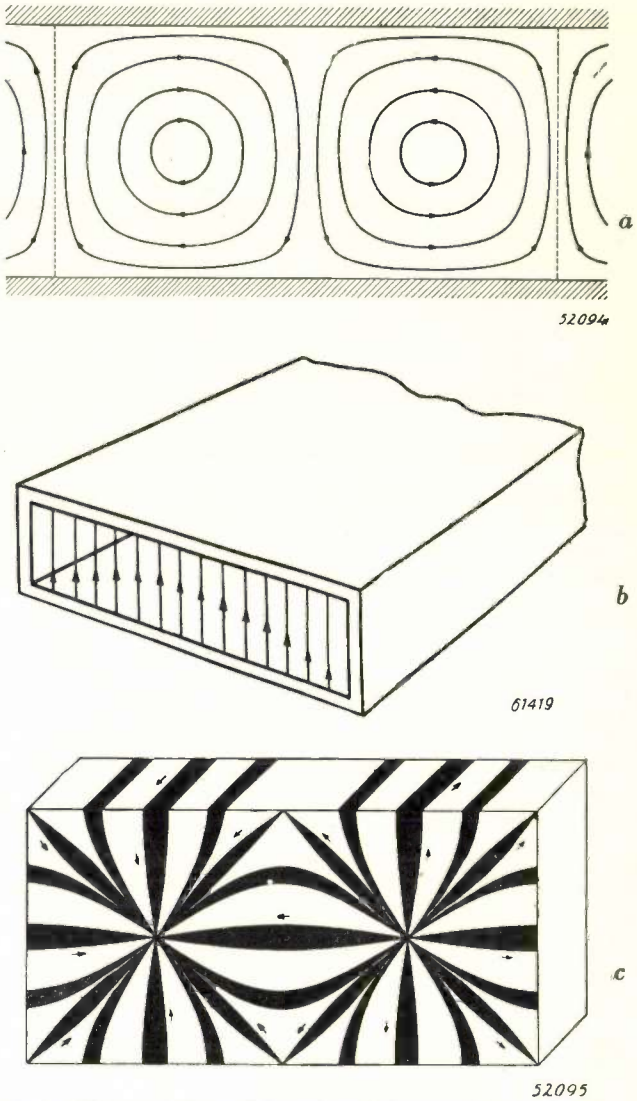


Fig. 4. *a*) Magnetic lines of force for a TE_{10} mode in a rectangular waveguide. This figure represents an "instantaneous photograph" of a cross section parallel to the wide sides, and thus also parallel to the direction of propagation of the wave. The wavelength at which the oscillation travels through the guide, λ_z, is equal to the distance between the two dotted vertical lines.
b) Electric field strength for a TE_{10} mode in a rectangular waveguide. The lines of force are all at right angles to the wide side.
c) Density distribution of the surface currents on the walls of a rectangular waveguide in which a TE_{10} mode is propagated. The width of the lines of current is proportional to the current density ("instantaneous photograph").

The occurrence of standing waves

When there is a discontinuity in a waveguide, for instance in the form of a sudden narrowing, then the wave will be reflected. The same may be the case at the end of the waveguide, for instance where it is coupled to the aerial, or where there is a bend in the guide.

In order to get an idea of the effect of such discontinuities we can avail ourselves of the similarity that exists between a waveguide in which, for instance, only the TE_{10} mode can be propagated,

and a transmission line. In the theory of transmission lines voltage and current are regarded as the fundamental quantities. In the case of a waveguide these quantities can no longer be unambiguously defined. One can, however, speak of reflections both in transmission lines and in waveguides. As a rule reflections occur in a transmission line when it is terminated with an impedance. The effect of a load impedance Z on the currents and voltages in a transmission line can be deduced as follows[3]).

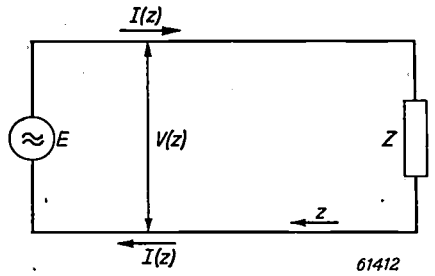


Fig. 5. Transmission line, with a load impedance Z, connected to the high-frequency voltage source E. At a distance z from the right-hand end the voltage and current are respectively $V(z)$ and $I(z)$.

As is known, the current and voltage at a certain point z in a transmission line (*fig. 5*) can be described by:

$$V(z) = \left\{Ae^{+j\gamma z} + Be^{-j\gamma z}\right\} e^{j\omega t},$$

$$I(z) = \left\{\frac{A}{\zeta} e^{+j\gamma z} - \frac{B}{\zeta} e^{-j\gamma z}\right\} e^{j\omega t},$$

where A and B are constants independent of z and t; γ is called the propagation constant. The quantity ζ is called the characteristic impedance of the system.

In the following we shall omit the time factor $e^{j\omega t}$, as not being essential for our considerations.

At the terminating impedance Z ($z = 0$) we have

$$V(0) = A + B \text{ and } I(0) = \frac{A}{\zeta} - \frac{B}{\zeta},$$

and therefore

$$Z = \frac{V(0)}{I(0)} = \frac{A+B}{A-B}\zeta \quad \ldots\ldots \quad (3)$$

Hence

$$\frac{B}{A} = \frac{Z-\zeta}{Z+\zeta} = f, \quad \ldots\ldots \quad (4)$$

and for the variation of the voltage we get

$$V(z) = A\left\{e^{+j\gamma z} + \frac{Z-\zeta}{Z+\zeta} e^{-j\gamma z}\right\} = A\left\{e^{+j\gamma z} + f e^{-j\gamma z}\right\}. \quad \ldots \quad (5)$$

The first term between brackets in this formula can be regarded as a wave travelling to the right and the second term as a wave travelling to the left, imagined as arising from reflection of the former wave at the impedance Z. The degree of reflection is determined by the equation

$$f = \frac{\frac{Z}{\zeta} - 1}{\frac{Z}{\zeta} + 1}, \quad \ldots\ldots\ldots \quad (6)$$

f being called the reflection coefficient, which is usually a complex quantity.

Owing to superposition of the waves travelling to the right and to the left, standing waves occur. At distances of a quarter of a wavelength there are voltage maxima alternating with voltage minima. These maxima and minima are given by the extreme values of (5) and are respectively

$$V_{max} = |A|\{1 + |f|\} \text{ and}$$

$$V_{min} = |A|\{1 - |f|\},$$

so that the ratio of the minimum to the maximum voltage, which indicates to what extent the resulting wave bears the character of a standing wave, is determined by the quantity

$$r = \frac{1-|f|}{1+|f|}. \quad \ldots\ldots \quad (7)$$

r is called the voltage standing-wave ratio (V.S.W.R.) for the impedance Z. By measuring the V.S.W.R., which can easily be done (see the article quoted in footnote [3])), the absolute value of the reflection factor can be determined. The argument of f, which is a measure of the phase shift resulting from the reflection, can be found from the distance between the point where the impedance Z is connected to the transmission line and the first voltage minimum, as already described in the article of footnote [3]).

Let us now revert to the rectangular waveguide. A discontinuity in the guide will result in a very complex wave pattern in its immediate vicinity, a pattern which can be regarded as a superposition of all the infinite number of modes of oscillation of the guide. Since only the TE_{10} waves are able to travel through the guide with small attenuation, at some distance in front of the discontinuity we shall find only the original wave and a reflected

[3]) See J. M. v. Hofweegen, Impedance measurements with a non-tuned Lecher system, Philips Techn. Rev. 8, 278-286, 1946.

wave. The superposition of these two waves again gives rise to standing waves. In analogy with the foregoing we can now introduce a reflection coefficient f, viz. the ratio of the amplitudes of the reflected and the original wave. Whereas in the case of the transmission line we had the ratio of two voltages to consider, here we have to reckon with the ratio of two electric fields, for instance the electrical field strengths of the waves in the forward and return directions.

In the case of transmission lines we found a reflection due to the presence of an impedance differing from the characteristic impedance ζ. With waveguides, too, an impedance can be ascribed to the discontinuity, by writing in analogy with (6):

$$f = \frac{\frac{Z}{\zeta} - 1}{\frac{Z}{\zeta} + 1}.$$

Just as current and voltage cannot be unambiguously determined, neither can the charateristic impedance ζ of a waveguide, except as the ratio Z/ζ in equation (6).

The magnitude of a standing wave in the waveguide can be expressed by the ratio of the minimum to the maximum amplitudes in the guide, thus by the voltage standing-wave ratio r, which, in analogy with (7), is related to $|f|$ according to the equation

$$r = \frac{1 - |f|}{1 + |f|}.$$

Under reflectionless conditions we find $f = 0$, thus $r = 1$, whilst for a perfectly reflecting termination $|f| = 1$, thus $r = 0$.

It is of great importance to keep the V.S.W.R. in a system as small as possible. When reflections occur, as for instance in a radar installation, these will react upon the oscillating magnetron and possibly disturb its stability. Moreover, although with a V.S.W.R. of say 0.6 the energy reflection from a discontinuity is still very small (about 6% = 0.5 db), the maximum power that can be sent through the guide without sparking occurring will already be 25% less than that when the guide is matched ($r = 1$).

A description will now be given of an arrangement used in the Philips Laboratories at Eindhoven for measuring the V.S.W.R.

Measuring the V.S.W.R.

When it is desired to match one element of a waveguide to another, for instance when the aerial of a radar system is to be coupled as efficiently as possible, it is necessary to measure the V.S.W.R. occurring as a result of reflection from that element, and to minimize this reflection. The latter can be done, for instance, by purposely introducing an obstacle, say an adjustable screw, which by its own reflections just neutralizes the reflections originally present. To confine our thoughts let us suppose that a detector has to be matched to a microwave installation in order to measure the strength of the signal in the waveguide. For this purpose we employ an arrangement as represented diagrammatically in *fig. 6*. The generator G produces a signal which is conducted along a number of parts, which will be described below, and finally reaches the detector K. There, as a rule, part of the signal will be reflected, with the result that standing waves arise in the whole of the guide. Placed in front of the detector is the "V.S.W.R. meter" S, consisting, for instance, of a short section of a waveguide in which the field can be scanned with a probe. This probe, not shown in the drawing, extracts a small part of the energy in the guide and passes it on to a measuring instrument connected to the probe. The voltage induced at the probe is proportional to the electric field at that point, so that by moving the probe along the guide it is possible to measure the voltage minimum and maximum due to the occurrence of the standing wave, and the ratio of these two voltages gives the V.S.W.R.

The probe can be made in the following way. First a slot is cut in the longitudinal direction of the waveguide down the centre of the long side;

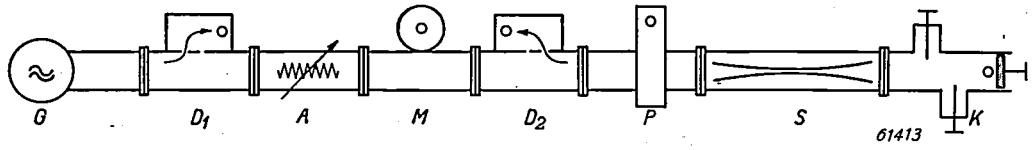


Fig. 6. Diagram of the lay-out for measuring voltage standing-wave ratios. G generator of the microwaves, D_1 and D_2 directional couplers, A attenuator, S slotted waveguide along which a probe (not shown) can be shifted, P fixed probe, M wavemeter, K detector which can be tuned and the voltage standing-wave ratio of which is to be measured. The small circles in the directional couplers and in the probe P represent crystal detectors with their matching elements.

this scarcely disturbs the system of surface currents, and thus the field, as may be seen from fig. 4c. Through this slot a thin wire can then be inserted a very short way into the guide (see *fig.* 7). This

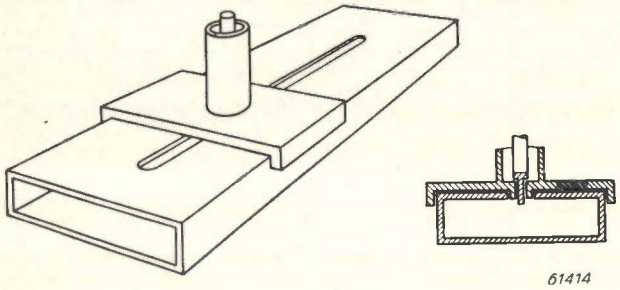


Fig. 7. Sketch of a probe that can be moved along the waveguide.

probe is then parallel to the electric lines of force and thus measures the electric field at the point where it is situated. The shorter the distance that the probe penetrates into the guide the less is the disturbance of the field, and it can usually be ignored. The probe is mounted on a carriage made to slide along the guide, care being taken to ensure that the probe is well centred in the slot and kept at the same depth of penetration into the guide. The latter requirement is not so easy to fulfil and a simpler method is therefore often applied, where a permanently fixed probe (P in fig. 6) and a waveguide of variable width are employed.

As we have already seen, an oscillation is propagated along the guide with a wavelength λ_z depending upon the width a of the guide (see formula (2) and fig. 3). This wavelength can therefore be varied by "squeezing" the guide and thus it is possible to change the number of wavelengths between the section of which the V.S.W.R. is to be measured and the fixed probe. This variation of the "electrical length" corresponds to a displacement of the field along the probe, so that by squeezing the guide, in which there is somewhere a fixed probe, we achieve the same object as when moving the probe along the field inside the guide in the manner described above.

A so-called squeeze section can be obtained by cutting a longitudinal slot down the centre of the two wide sides and placing midway over it a clamp with which the guide can be constricted a few millimetres. In *fig.* 8 two types of a squeeze section are shown, one for 3 cm waves and the other for 10 cm waves. The guide for 10 cm waves is fitted with a dial gauge for measuring the constriction of the squeeze section.

It is advantageous to squeeze the guide periodically. In the Philips Laboratories a method is employed whereby the guide is squeezed once in about every two seconds by means of two cams driven by an electromotor. The voltage on the probe will vary in rhythm with the constrictions, and the ratio of the lowest to the highest voltage indicated by a meter which has little inertia is a measure for the V.S.W.R.

Here the "probe" does not consist of a wire but is formed by a so-called window communicating between the main guide and a section of waveguide mounted on it crosswise and containing a detector. After amplification the voltage supplied by the detector can be measured; we shall revert to this later on.

A V.S.W.R. meter such as described here is only suitable for measuring small standing-wave ratios. It often happens that the matching of a variable waveguide section is at first so bad that the V.S.W.R. has a very high value. For quickly matching the section in such a case it is advisable

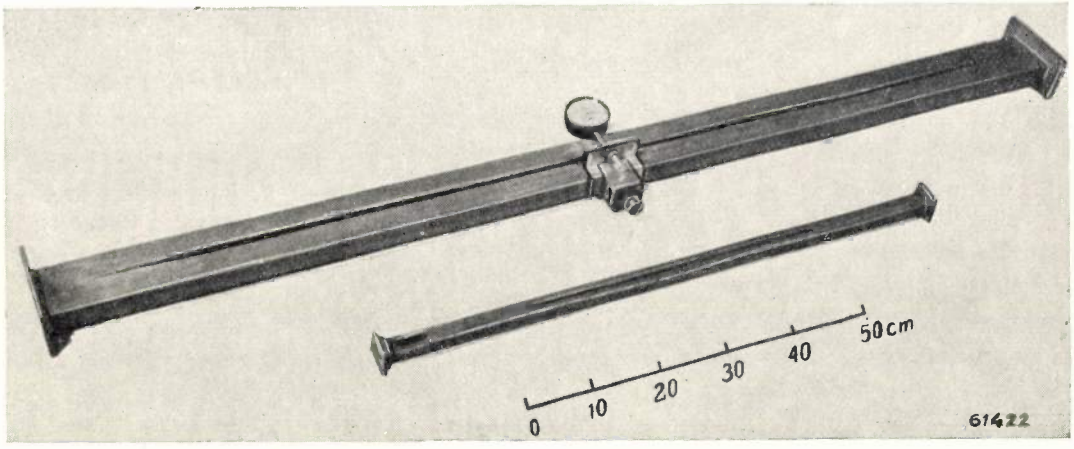


Fig. 8. Two types of a squeeze section . The large section (for 10 cm waves) is fitted with a dial gauge for calibrating the electrical length of the waveguide as a function of the width of the slit.

to have available a coarser indicator of the V.S.W.R. This is to be found in what is called the directional coupler, by means of which the strength of the reflected signal can be measured. A directional coupler consists of a branch waveguide connected to the main waveguide by windows and in which a wave is propagated which is a certain fraction of a wave travelling in one direction along the main waveguide. A diagrammatic representation of such a directional coupler is given in *fig. 9* [4]). The

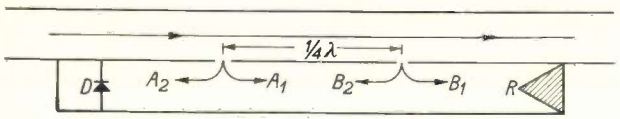


Fig. 9. Basic principle of a directional coupler. The two windows in the common, narrow wall of the waveguides are spaced at a distance of $^1/_4\ \lambda_z$. A wave travelling to the right sets up through each window a secondary wave travelling to the left and to the right. The wave B_2 follows a path half a wavelength longer than that of the wave A_2. Thus the two waves neutralize each other. A_1 and B_1 on the other hand reinforce each other, so that in the secondary waveguide there is exclusively a wave travelling to the right. Similarly, a wave travelling to the left in the main waveguide will give rise exclusively to a wave travelling to the left in the secondary guide. The wave travelling to the right can be absorbed by a non-reflecting termination R. Thus the detector D receives only a signal from the wave travelling to the left in the main waveguide.

two identical windows in the common, narrow, side wall of the two waveguides are spaced a quarter of a wavelength apart ($^1/_4\ \lambda_z$). A wave travelling to the right through the main guide gives rise to a secondary wave passing through each of the windows into the secondary guide and travelling to the right and to the left. The waves A_1 and B_1 reinforce each other; the waves A_2 and B_2 however neutralize each other through interference, since the wave B_2 follows a path half a wavelength longer than that followed by the wave A_2. For proper neutralizing of the waves A_2 and B_2 it is necessary that the coupling should be weak, or in other words: the attenuation of the wave in the main guide caused by a window must be negligible, so that this wave can pass both the windows with the same strength and thus the waves A_2 and B_2 formed in the secondary guide are of the same intensity.

In this way a wave travelling to the right in the main guide results exclusively in a wave travelling to the right in the secondary guide; similarly, a wave travelling to the left in the main guide produces in the secondary guide exclusively a wave travelling to the left. The right-hand end of the secondary guide can now be closed with a non-reflecting termination (for instance with a wedge of dissipative material) and on the left-hand end of the guide a detector can be mounted.

This detector will then pick up only a signal from the wave travelling towards the left in the main guide and a wave travelling to the right in this guide will not reach the detector. In this way it is therefore possible to detect at once the presence of a reflected wave.

What it amounts to in practice is that, measuring with the directional coupler, we can make the matching of a waveguide section such that the V.S.W.R. for that element exceeds 0.95, which corresponds to an energy reflection of 0.06%. Measuring with the squeeze section we can then carry out the accurate adjustment, since it is possible to measure a V.S.W.R. of 0.99, corresponding to an energy reflection of only 0.0025%.

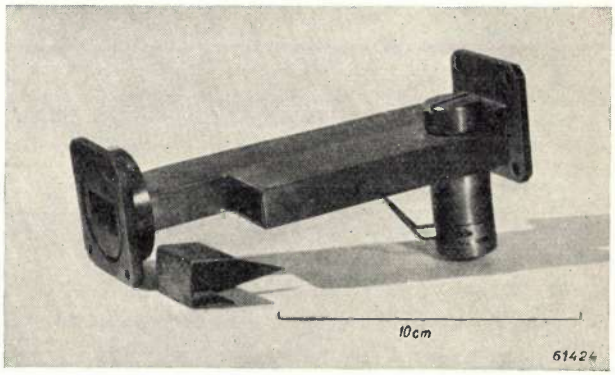


Fig. 10. Photograph of a directional coupler for a 3 cm waveguide. The wedge acting as a matched termination of one of the ends has been removed from the secondary waveguide.

A photograph of a directional coupler for a 3 cm waveguide is given in *fig. 10*. By way of illustration the wedge for the non-reflecting termination of the section is shown beside the coupler.

Fig. 11 is a photograph taken of the whole of the set-up described here.

Accessories for the measuring device

We shall now describe a few accessories (some of which have already been mentioned) which may be useful for a measuring device as described.

A generator is needed for generating the microwaves used in carrying out the measurements. For this one usually employs a velocity-modulation valve (reflex klystron), which is capable of supplying a power of some milliwatts up to some tens of milliwatts. This valve oscillates only when the value of the reflector voltage lies within certain

[4]) See K. S. Knol and G. Diemer, A model for studying electromagnetic waves in rectangular wave guides, Philips Techn. Rev. 11, 156-163, 1949 (No. 5).

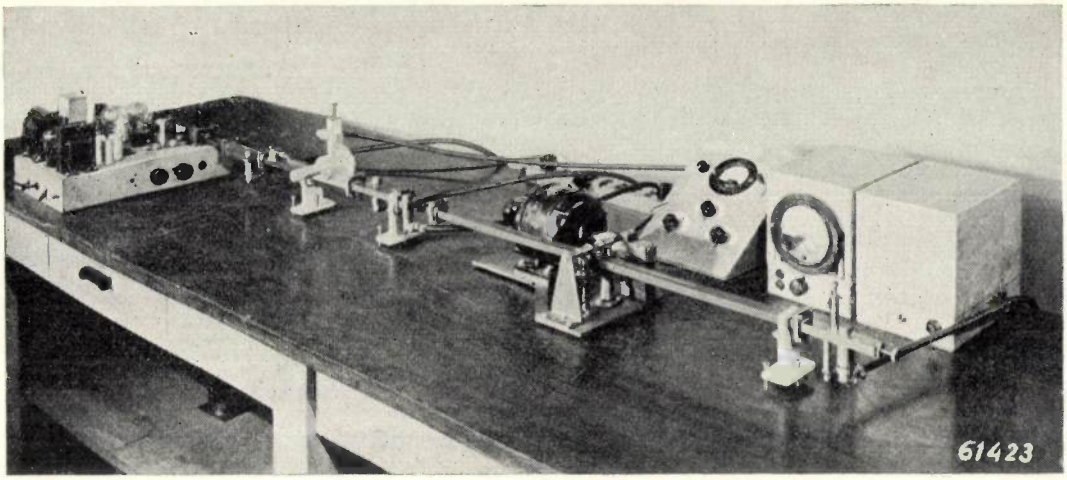


Fig. 11. Photograph of the complete device for measuring standing waves, as employed in the Philips Laboratories.

ranges. The frequency of the energy generated can be varied over a band width of about 10% by mechanical tuning of the cavity resonator of the valve, and further over a more limited range by varying the reflector voltage within an oscillating range. For our measurements an A.F. amplitude modulation of the signal is desired. The voltage picked up by the fixed probe in the squeeze section can then be amplified, after detection, with the aid of an A.F. amplifier. Although the amplitude of the oscillations greatly depends upon the reflector voltage, the method of modulating the amplitude of the signal by sinusoidally varying this voltage is not suitable, because then frequency modulation occurs at the same time. It is possible, however, to get a pure amplitude modulation by applying to the reflector square waves of such an amplitude that the valve is alternately oscillating and not oscillating. This is the method employed with the measuring device described here.

The microwaves are conducted from the oscillator to the waveguide via a short coaxial cable (*fig. 12*). At one end of this cable the inner conductor is bent into a small loop, which is connected to the outer conductor, and this loop is inserted in the cavity resonator of the reflex klystron. At the other end of the cable the inner conductor is extended into the wave guide parallel to the electric lines of force, like an aerial, whilst the outer conductor is connected to the wall of the waveguide. The guide is closed at one end with a piston, in such a way that maximum power is obtained in the desired direction.

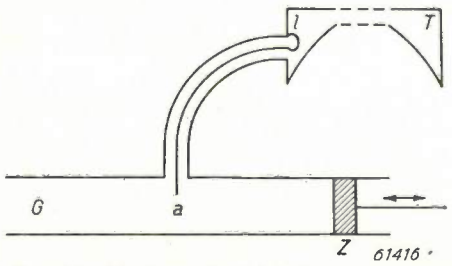


Fig. 12. Coaxial cable through which the microwaves generated by the reflex klystron are conducted to the waveguide. The small loop *l*, inserted in the cavity resonator of the valve, takes up the energy; the "aerial" *a* is the inner conductor of the coaxial cable extending into the waveguide *G* parallel to the electric lines of force. Here the measuring apparatus is to the left of the aerial; the waveguide, however, has been extended somewhat to the right to allow of the terminating of the guide at that end with a piston *Z* which can be adjusted so that the maximum energy is obtained in the desired direction.

As the first element behind the generator there is usually a variable strip attenuator, by means of which the signal amplitude can be adjusted while at the same time the generator is decoupled from the system; in the case of quantitative measurements care has to be taken that the strength and the frequency of the signal are not affected by the generator load. The strip attenuator (*fig. 13*) is made in the form of a small sheet of dissipative material placed in the waveguide, the attenuation being brought about by eddy currents set up in the absorptive strip. Here again use is made of the possibility of cutting a longitudinal slot down the centre of the wide side of the guide without thereby disturbing the field inside the guide. The attenuation is controlled by the depth of penetration of the strip of dissipative material in the guide. The perimeter of the strip is rounded off so as to minimize reflections.

Mention has several times been made of a detector, which rectifies the H.F. voltages so as to be able to measure the energy. As such a detector a diode [5]) can still be used for waves of 10 cm and

[5]) M. J. O. Strutt and K. S. Knol, A diode for the measurement of voltages on decimetre waves, Philips Techn. Rev. 7, 124-128, 1942.

longer, but for shorter waves the diode is unsuitable on account of transit-time effects, and then a crystal detector has to be used. In principle a crystal detector does not differ from the crystals

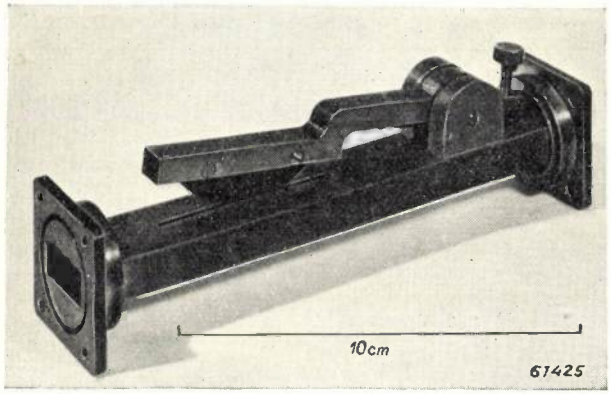


Fig. 13. Attenuator consisting of a small sheet of dissipative material which can be inserted in the waveguide through a longitudinal slot in the wide side.

that were formerly employed in radio engineering; recent investigations in America and England have made it possible to stabilize crystals, so that reliable and reproducible results can now be obtained. The crystal detector is highly suitable for microwaves owing to its small dimensions and small shunt capacitance.

The crystal is mounted in a cartridge (a ceramic tube with metal end caps) which is secured in a crystal holder consisting of a waveguide section of the same dimensions as the guide in which the power is to be measured (see *fig. 14*); the guide is terminated by a short-circuiting piston and fitted with some fixed or adjustable tuning elements, with which the matching can be so adjusted that all incident energy is absorbed in the crystal and thus no reflection takes place. The cartridge is placed with its longitudinal axis parallel to the electric field. The matching of a crystal holder has been taken as example in the description of the measuring of the V.S.W.R. in the preceding section of this article.

Another possible means of detection lies in the employment of elements the electrical resistance of which is dependent upon the temperature. Such an element forms part of a Wheatstone bridge circuit, with which the variation in resistance due to the heating of the element by the absorbed microwave energy is measured. Use can be made either of thin wires with a positive temperature coefficient of the resistance, or of semi-conductorss having a high negative temperature coefficient [6]) and usually of the "pin head" type enclosed in a glass envelope; the whole is placed in a holder with the leads running parallel to the high-frequency electric field. The advantage of such "bolometers" is that they can be directly calibrated with direct current and thus form direct wattmeters. On the other hand, however, they are less sensitive than crystal detectors.

The voltage supplied by the detector in the fixed probe used for measuring the V.S.W.R. is applied to the input of a linear selective amplifier, tuned to the frequency of the above-mentioned amplitude modulation with rectangular pulses of the oscillator signal. The output signal from the amplifier is rectified and passed to a meter provided with a scale giving direct readings of the V.S.W.R.

In addition to the directional coupler D_2, serving for the rough matching of waveguide sections with a large V.S.W.R., there is a second directional coupler (D_1 in fig. 6). This is coupled with the generator signal and serves for adjusting the generator output to the maximum with the aid of the voltages applied to the tube and further with the piston mentioned when dealing with the construction according to fig. 12.

Finally the system includes a frequency meter or wave meter for ascertaining the correct wavelength of the signal generated by the reflex klystron. The wavemeter contains, as resonating element, an adjustable cavity resonator or some other tuned system coupled to the waveguide. When the system is so tuned that resonance occurs at the frequency of the microwaves in the waveguide then the energy taken up in the system will be the maximum. There are two ways of determining this:

a) The tuning can be determined at which the

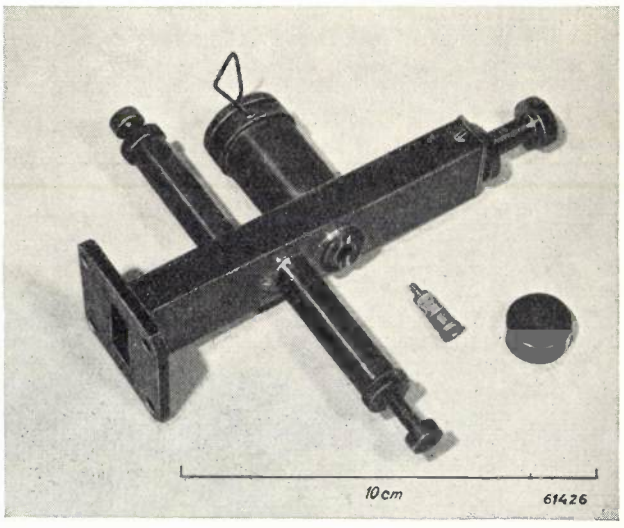


Fig. 14. Detector; the two lateral tubes fitted with a tuning screw are adjusting elements; the screw in the right-hand end of the waveguide serves for adjusting its terminating piston. The cartridge containing the crystal has been taken out of the detector.

[6]) See E. J. W. Verwey, P. W. Haayman and F. C. Romeyn, Semi-conductors with large negative temperature coefficient of resistance, Philips Techn. Rev. **9**, 239-248, 1947.

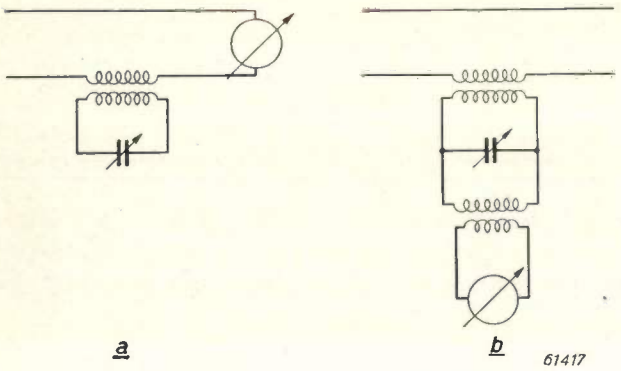


Fig. 15. Two methods for measuring the frequency of the microwaves with the aid of a wavemeter. *a*) Reaction-type wavemeter; the tuning is determined at which the energy loss in the waveguide due to absorption in the resonating wavemeter reaches the maximum. *b*) Transmission-type wavemeter; the tuning is determined at which the energy transmitted by the resonating system to the meter reaches the maximum.

loss in power suffered by the wave travelling in the waveguide as a result of the presence of the wavemeter reaches the maximum; for instance by measuring the voltage induced in the fixed probe of the V.S.W.R. meter (in the case of constant width of the waveguide) as a function of the adjustment of the wavemeter. When this voltage is the minimum the wavemeter resonates and from its adjustment one can determine the frequency (reaction-type wavemeter).

b) It can be directly determined when the power taken up by the wavemeter is greatest, with the aid of a detector and a voltmeter connected to the wavemeter (transmission-type wavemeter; the power measured is now the the maximum when resonance occurs).

These two methods of measuring are represented diagrammatically in *figs. 15a* and *15b*.

It depends upon the wavelength of the microwaves what kind of tuned system is used. For 3 cm waves an adjustable cavity resonator is mostly used (*fig. 16a*); one of the end plates of this resonator is adjustable with a micrometer screw.

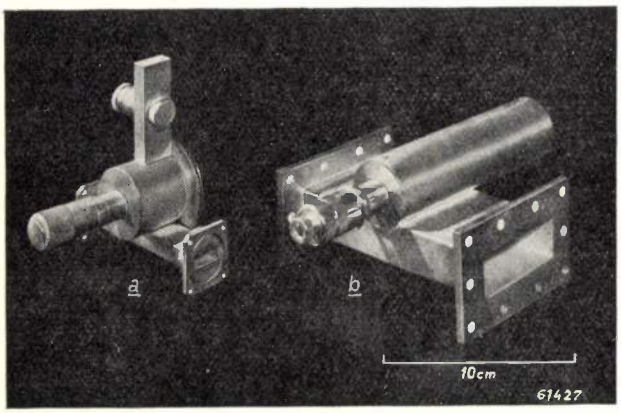


Fig. 16. *a*) Wavemeter in the form of a variable cavity resonator for 3 cm waves. One of the end plates can be moved with a micrometer screw. *b*) Wavemeter for 10 cm waves; this is in the form of a transmission line which can be tuned. Both wavemeters are coupled to the waveguide by means of windows.

For longer waves (10 cm) usually a tuned coaxial transmission line is employed (fig. 16*b* and, diagrammatically, *fig. 17*). At resonance the inner conductor is $^3/_4$ wavelength long; the sliding contact on the rod is $^1/_4$ wavelength distant from the shortcircuited end, for the average frequency for which the meter is suitable, so that it always lies near a current node and the contact does not have to answer stringent requirements.

The coupling with the waveguide is usually brought about by mounting the cavity resonator or transmission line direct on the waveguide and allowing the electromagnetic field to enter the wavemeter via a window, as is to be seen in figs. 16*a*, 16*b* and 17.

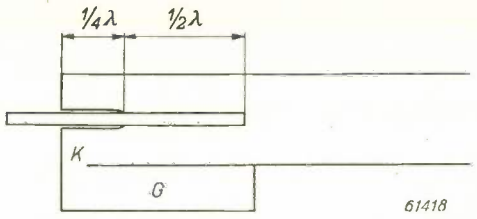


Fig. 17. Cross-section of the coaxial transmission line of fig. 16*b*. At resonance the inner conductor is $^3/_4$ wavelength long, the sliding contact being at $^1/_4$ wavelength from the short-circuited end of the outer conductor. The waveguide *G* (the longitudinal direction of which is at right angles to the plane of the drawing) is coupled to the wavemeter by the window *K*. The transmission line is open-ended at the right; by suitable choice of the dimensions of the system the waves are prevented from being propagated in the right-hand part of the cylinder, where there is no inner conductor.

Summary. After explaining the necessity of employing waveguides in microwave installations for high powers, consideration is given to the problems arising from the use of these waveguides. Discontinuities, bends and faulty matchings in the waveguide cause reflection of the electromagnetic waves; these reflections have to be avoided as far as possible, and it is therefore necessary to be able to measure the reflective properties of a waveguide section. The effect of discontinuities in the waveguide is explained by comparison with a transmission line. Owing to reflection a standing wave is formed; the absolute value of the reflection coefficient is determined by measuring the voltage standing-wave ratio (V.S.W.R.) in the wave-guide.

For measuring the V.S.W.R. a waveguide section is used in which the electric field can be determined from point to point with the aid of a probe. In addition to a probe sliding along the guide it is also possible to employ a fixed probe; by "squeezing" the guide one can change the wavelength of the oscillation in the waveguide and thus cause the field to be displaced past the probe. Such a V.S.W.R. meter is only suitable for measuring fairly small standing-wave ratios. To determine roughly the correct matching of a waveguide section, use is made of a directional coupler, with which the strength of a signal travelling in one direction can be measured.

Finally the accessories of the measuring device are dealt with separately; brief descriptions are given of the construction of the generator for the microwaves, of the attenuator, of the crystal detector rectifying the high-frequency signal, and of the wavemeter for measuring the frequency of the oscillations.

ON ELECTRICAL SHAVING

by A. Th. van URK.

672.715.3-83.004.13

While shaving, one may well ponder about such questions as to how exactly the hairs of the beard grow, how fast they grow in a day, how thick they are and how many cover a square centimetre of the skin. It is the knowledge of such matters as these, supplemented by experiments relating particularly to electrical shaving, that forms the basis for the rational construction of an electric dry-shaver.

Skin and hair

The human skin, a cross-sectional diagram of which is given in *fig. 1*, consists of three layers: the outer skin, or epidermis, the leathery skin, or corium, and the lower skin, or subcutis. As far as shaving is concerned it is of course mainly the outer skin that we have to deal with. This is a mucous membrane built up in a number of layers, with the cells gradually migrating to the surface, where they become flattened out parallel to the surface

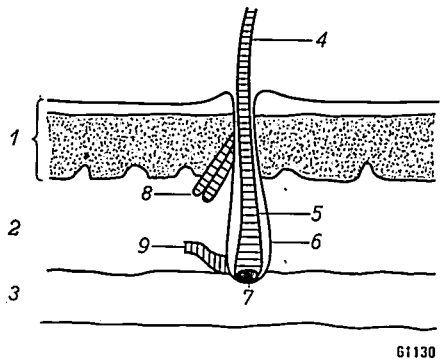


Fig. 1. Cross section of a part of the human skin, with a diagrammatic representation of a hair. *1* epidermis, *2* corium, *3* subcutis, *4* hair shank, *5* root of the hair, *6* hair sac, *7* hair papilla, *8* sebaceous gland, *9* hair muscle.

and ultimately are cast off in the form of horny scales. Consequently there are two main layers in the outer skin. The lower one is formed of living cells which continuously divide themselves, and it is therefore called the germinative layer, or stratum germinativum. The other is a surface layer, called the horny skin, or stratum corneum, which acts as a protective covering; it consists of cells that are no longer capable of dividing themselves.

There is a very intimate connection between the outer skin and the corium, on account of the finger-shaped spurs, called papillæ, extending from the corium into the epidermis. These papillæ contain capillary blood vessels, and as a consequence a cut extending down to the level between the epidermis and the corium is apt to cause bleeding, as is often the experience when using a razor. The thickness of the epidermis on the cheek of an adult is between 90 and 120 microns, i.e. about one-tenth of a millimetre. The papillæ do not extend more than about 8 microns into the epidermis of the cheek.

The hairs of the beard of an adult male are formed in the skin. The part protruding from the skin is the hair shank, the scapus pili, whilst the part hidden in the skin is called the radix pili, or the hair root. As shown in fig. 1, this root is contained in a sac or follicle, called the folliculus pili, which is embedded in the mucous membrane of the epidermis and in the corium and sometimes extends down into the subcutis. At the bottom of the hair sac is the hair papilla, bulbus pili, a spherical protuberance out of which the hair grows. Finally, at the side of the hair sac are sebaceous glands feeding the hair with fat, and below these is a muscle, the musculus arrector pili, which enables the hairs to erect themselves at times, or, as the saying goes, to stand on end.

It has been found that the daily growth of the hairs forming the beard varies between 0.2 and 0.6 mm, averaging 0.4 mm.

Density of the beard

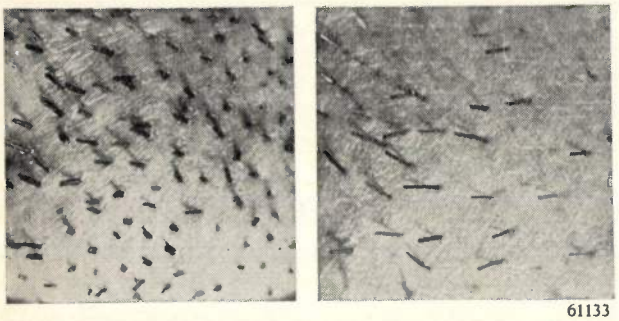
To form an idea of the density of the beards of various people, photographs have been taken of parts of the beard of a number of test persons and from the greatly magnified photographs the numbers of hairs per square centimetre have been counted. As is only to be expected, these numbers differ considerably, as is evident from the following figures:

on the chin: 40 - 100 - 120 - 100 - 110 - 110 - 70 - 80 - 90 - 100 - 70 - 90 - 90 - 70, averaging 90 hairs per cm^2;

on the cheek: 45 - 60 - 70 - 85 - 70 - 60 - 60 - 80 - 35 - 80 - 50 - 55 - 35 - 40, averaging 60 hairs per cm^2.

In a beard there is therefore on an average no more than one hair per mm^2 (by way of comparison, the density of the hair on top of the head can be

taken as 3 hairs per mm². According to these results [1]) and taking into account the area covered by the growth of beard, a normal beard has about 13,000 hairs.



61133

Fig. 2. Magnified photographs of part of the beards of two test persons. Left: part of the chin, 110 hairs per cm²: right: part of the cheek, 35 hairs per cm².

Two of the photographs mentioned above are reproduced in *fig. 2*.

Thickness of the hairs

For the construction of a shaving apparatus it is of essential importance to know the maximum thickness of the beard hairs: it is obvious that the minimum width of the slots or holes of the apparatus into which the hairs have to be drawn before they are cut has to be greater than the maximum thickness of hair likely to be met with.

This has been investigated by taking 31 "shavings" and measuring the thickest hairs found in them. The results are graphically represented in *fig. 3*. It appeared that the maximum thickness most commonly occurring was 0.14 mm. In about 90% of the number of shavings examined the maximum thickness of the hairs in any one of them proved to be less than 0.18 mm.

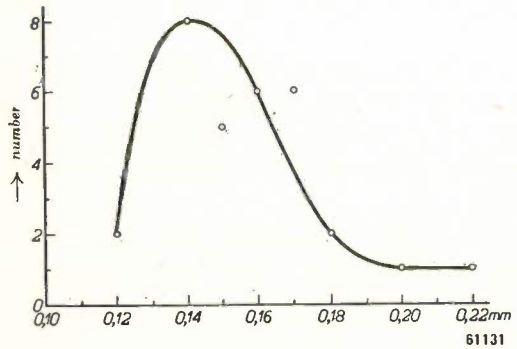


Fig. 3. The greatest thickness of the hairs contained in each of 31 "shavings" was determined. For each value it has been plotted how many times it occurred as maximum value. It appears that in about 90% of the shavings examined the maximum thickness of hair is less than 0.18 mm.

[1]) The figure of 40 hairs per cm² sometimes found in the literature on the subject is apparently based on inaccurate data.

It is also possible to get some idea of the average thickness of beard hairs. The average weight of 24 "Philishave" shavings from a beard growth of 24 hours was found to be 28 mg. Taking the numbers of hairs in a beard at 13,000, as previously calculated, with an average length of 0.4 mm, and reckoning the specific gravity of hair as being approximately 1, the average thickness of the hairs in an average beard works out at about 0.07 mm.

The hair "streams" in a beard

The roots of practically all hairs are planted obliquely in the skin. From this, and the fact that neighbouring hairs stand at an angle to the surface of the skin in almost the same direction, it follows that the hairs of the beard grow in different directions, depending on what part of the face they are growing. We therefore speak of hair streams. In these streams the directions of the hairs are nearly always at a small angle to each other, and thus, instead of being parallel,

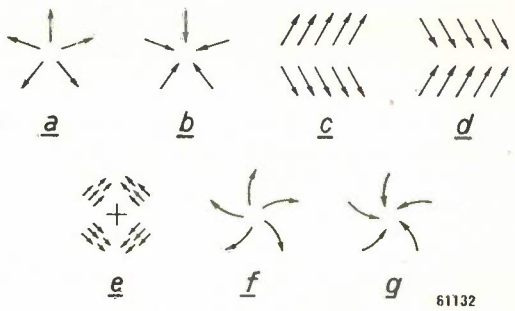


Fig. 4. Various examples of "hair streams" occurring in the human beard: *a*) stream with a centre of radiation, *b*) stream with a centre of attraction, *c*) diverging and *d*) converging stream, *e*) cross-shaped growth, *f*) diverging and *g*) converging eddy stream.

converge or diverge. Indicating the directions of the "stream lines" by arrows, we get diagrams greatly differing in shape, some examples of which are given in *fig. 4*.

These hair streams, which are characteristic for each individual, have to be taken into account when shaving, because shaving against the stream is more effective than shaving with the stream. When shaving electrically one instinctively feels how the direction of movement of the shaving head has to be changed, according to what part of the face is being shaved; sometimes the shaving head has to be moved in a straight line and sometimes it has to be rotated.

Construction of the "Philishave" dry-shaver

Having now said something about the growth of the hairs of the beard, we shall consider how these facts can be taken into account in the construction of a dry-shaver, in this case the "Philishave" dry-shaver [2]).

This apparatus consists of a stationary part, the shaving cap, which is pressed against the skin and has a large number of slots into which the hairs are guided, and a moving part that cuts off the hairs. The latter part has a rotating motion, and not an oscillating motion as in the case of a pair of clippers. It consists of a six-fold milling cutter rotating at a high speed and pressed gently by a spring against the inner face of the stationary cap.

Compared with the oscillating movement, the rotating motion has the advantage that a very much higher cutting speed can be reached, this being helped by the presence of the six-fold cutter. The cap has been given such a diameter that the whole of its surface can always be in contact with the skin. A secondary advantage, but not unimportant from the hygienic point of view, of the construction of this apparatus is that the hairs which have been cut off are collected in the cap instead of being allowed to fly about. This apparatus and some of its component parts are illustrated in *fig. 5*.

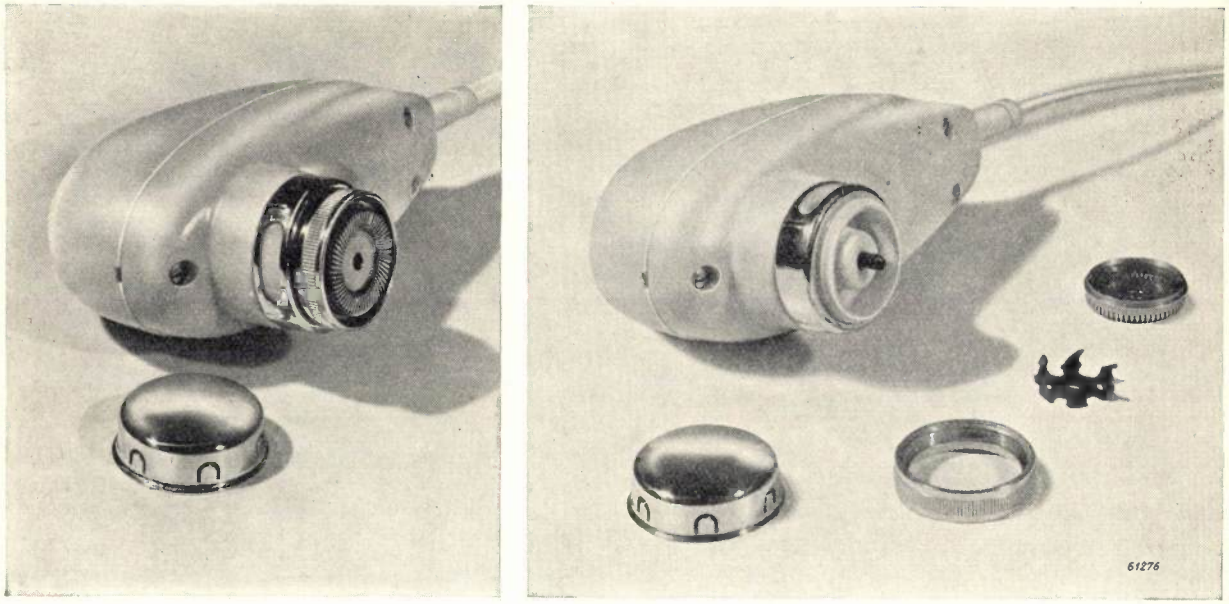


Fig. 5. On the left the "Philishave" dry-shaver and the guard cap. On the right the same but with the shaving cap and cutter detached.

What one desires with a shaving apparatus is to be able to shave quickly and cleanly, without any risk of injuring the skin. In this respect the width of the slots and the thickness of the cap where the cutter is rotating are particularly important.

Let us first consider the width of the slots. We have seen that the average thickness of the beard hairs is 0.07 mm, but to be able to catch also the thickest hairs the slots must in any case be wider than 0.18 mm. The wider the slot the greater the chance of catching the hairs, since roughly speaking this chance is proportional to the difference between the width of the slot and the thickness of the hair. For quick shaving, therefore, a fairly wide slot is desired. On the other hand, however, if the slot is too wide there is a risk of the skin bulging out too far into the slots — a point which will be further dealt with presently — and thus suffering damage. A width of about $^1/_4$ mm appears to be a good compromise.

The slots on the inside of the shaving cap, where the cutter describes its circular path, are extended to the edge of the cap (see fig. 5) and join up with the guiding grooves in the cylindrical wall of the cap. In order to ascertain how and where the hairs enter the slots, film photographs were taken of a beard rubbing over a shaving cap. When the test

[2]) A description of this apparatus was given by A. Horowitz, A. van Dam and W. H. van der Mei, "The dry-shaving apparatus "Philishave", Philips Techn. Rev. 4, 350-354, 1939. In the design described in that article a three-fold milling cutter was used, but the more recent designs have a six-fold cutter.

person rubs his cheek over the cap at a rate of say 10 cm per second and the slots are 0.25 mm wide, a hair will pass across the slot in about 1/400th of a second. A normal film camera taking 24 pictures per second is therefore not suitable for showing how a hair is caught up in the slot. Consequently a camera with a very high speed had to be used, taking up to 3000 pictures per second. A magnified cut from one of the films is reproduced in *fig. 6*. This investigation showed, inter alia, that long hairs enter the slots mainly via the guiding grooves, while shorter ones may appear anywhere in the slot.

The second important factor mentioned above was the thickness of the shaving cap. At first sight one would be inclined to say that the thinner the cap the better it is — provided it remains sufficiently rigid — because with a thin cap the hairs will be cut off closer to the skin than is the case with a thicker one. Actually, however, what takes place is that the skin, which is more or less supple, bulges out into the slots and thus helps in cutting the hairs off short. But since it is not desirable for the skin to penetrate too far into the slots, as we have already seen, on that account alone a limit is set to the thinness of the shaving cap, apart from the question of rigidity.

An investigation has also been made to see in how far the skin tends to bulge out into the slots. The device used is illustrated in *fig. 7*. From the results obtained it appeared that the extent of penetration (thus the suppleness of the skin)

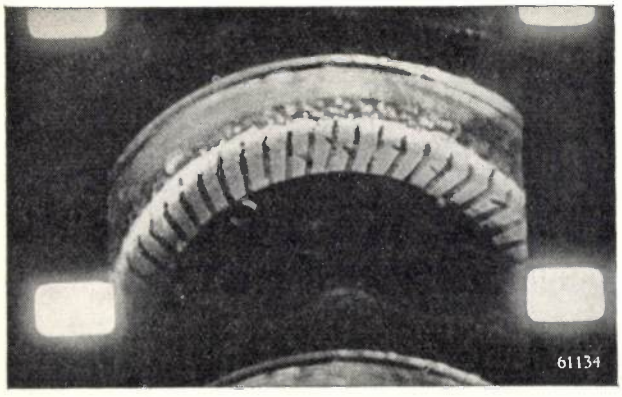


Fig. 6. One of the pictures from a cinematographic film showing how the hairs of a beard are caught up in the shaving cap of a "Philishave" dry-shaver.

not only differs considerably as between one person and another but also shows rather great differences for the same person between the various parts of the face (cheek, chin, neck). Naturally the extent of penetration also depends for a great deal upon the force with which the shaving cap is pressed

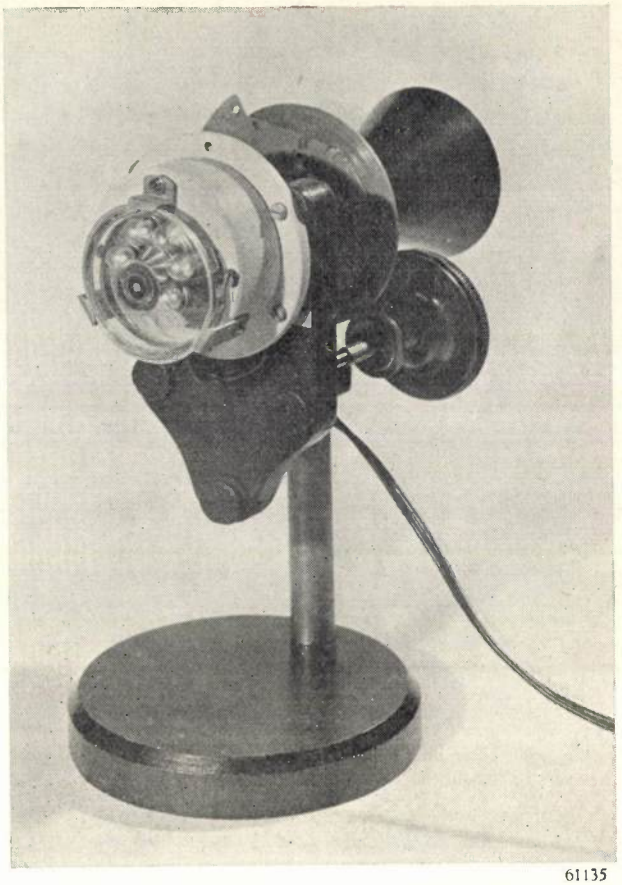
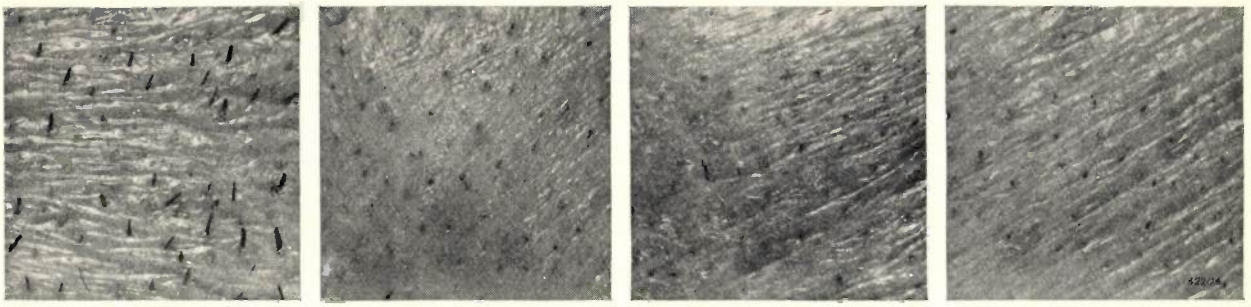


Fig. 7. Microscope set-up used for measuring the extent to which the skin bulges out when shaving with a dry-shaver. The shaving cap is centred in the round glass cap at the front of the apparatus, and is illuminated from the inside by a number of small lamps placed in a circle round the objective. The microscope eye-piece is inside the cone on the right-hand side of the apparatus. The observer focuses the microscope onto the inner face of the shaving cap along which the cutters run, and then onto the skin; the displacement gives a measure for the bulging of the skin.

against the skin, but when using the apparatus one very soon acquires the "feeling" to know on what parts of the face the shaving head has to be pressed down more or less heavily in order to get a clean shave. In the light of these considerations, with a slot width of $^1/_4$ mm about 0.1 mm is found to be the most suitable thickness for the material of the shaving cap.

Finally, in *fig. 8* four enlarged photographs are given of part of the beard of a test person shaved in three different ways, viz. with a safety razor, by a barber with an ordinary razor and with a "Philishave" dry-shaver. In all three cases the face is sufficiently clean-shaven (disregarding the single hair accidentally left by the barber, photo *c*). In the case of photo *b* it may aptly be said that the "safety" razor has shaved off the beard "with hide and hair", whereas photo *d* shows that the "Philishave" dry-shaver — thanks to the carefully cal-



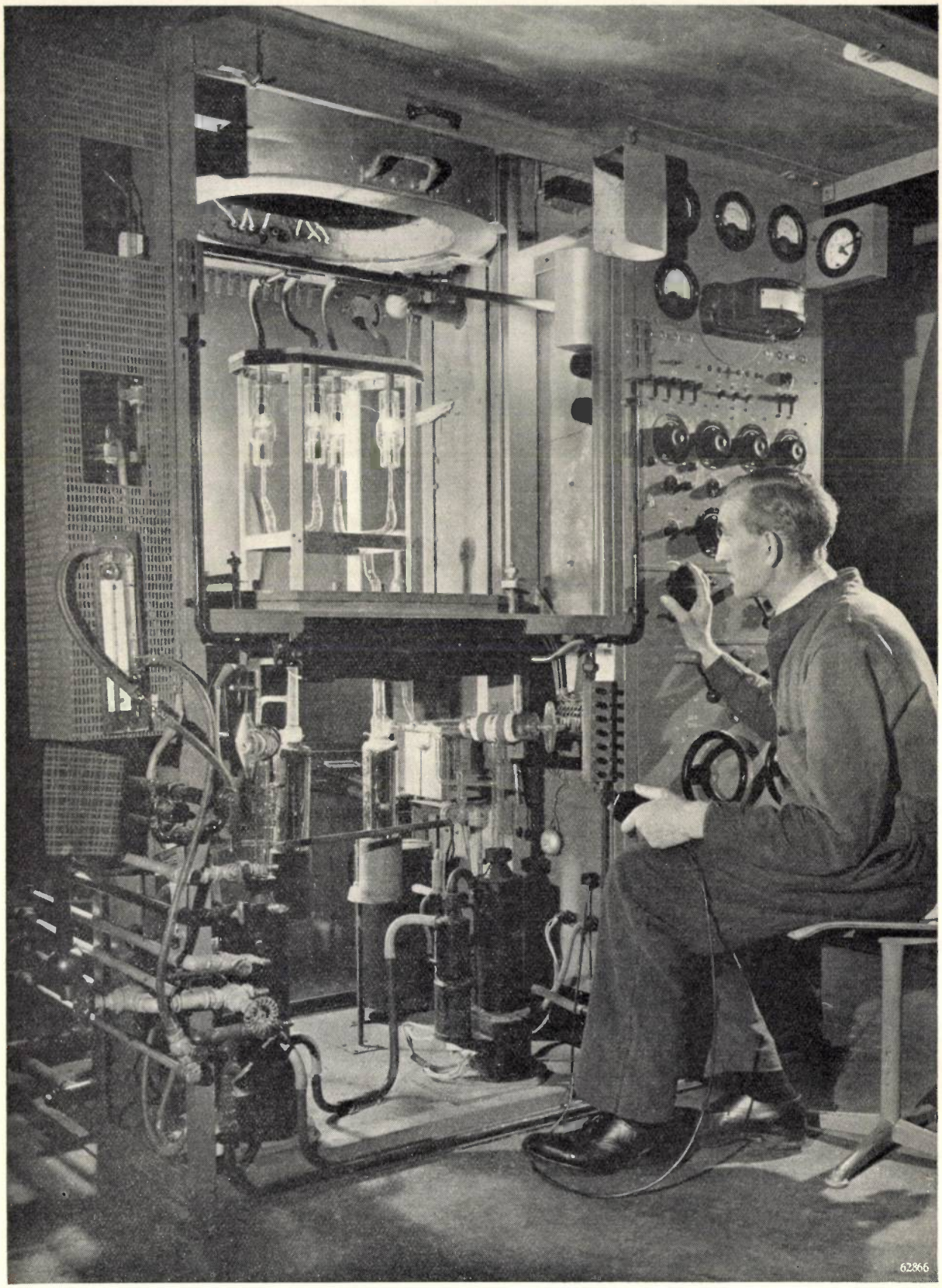
a *b* *c* *d*

Fig. 8. The same part of the beard of a test person, 8 times enlarged, *a*) before shaving, *b*) after shaving with a safety razor, *c*) after being shaved by a barber, *d*) after shaving with a "Philishave" dry-shaver.

culated dimensioning of the shaving head — has left the original structure of the skin practically untouched, better even than the professional barber with his razor.

Summary. From data obtained from the literature on the subject and from his own observations the writer gives a brief account of the growth of the hairs of the beard. The number of hairs in a normal beard amounts to about 13,000 and the growth per day is about 0.4 mm. The maximum thickness of the hairs has been found to be 0.18 mm and the average thickness 0.07 mm. For quick and clean shaving with the "Philishave" dry-shaver the width of the slots and the thickness of the material of the shaving cap are of primary importance. It is described how, with the aid of a high-speed camera taking up to 3000 pictures per second, a study has been made of the manner in which the hairs of the beard are caught up in the shaving cap. A brief description is also given of the means employed for measuring the extent to which the skin bulges out into the slots when shaving.

PUMPING OF X-RAY TUBES



With this apparatus four X-ray tubes (diagnostic tubes for 110 kV, of the "Statix" type) are pumped simultaneously. The elliptical oven seen raised above the tubes can be lowered over them for heating the tubes to the temperature required for degassing.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1892: J. Haantjes: A self-oscillating line-deflection circuit (Bull. S.E.V. **40**, 633-635, 1949, No. 1).

Description of a self-oscillating circuit in which, apart from a short fly-back time, the voltage is a linear function of time. Schematically the circuit may be represented by a self-inductance L and a capacitance C connected in parallel to a battery via a switch. If the switch is periodically opened for a short time, equal to the time of one half oscillation of the circuit, oscillations of the desired properties will occur. In the article it is shown how the switch may be realised by a combination of a diode and a triode. By a slight modification the internal resistance of the switch may be lowered. Some other alterations are described, which allow the circuit to be synchronized, on the control grid, with negative pulses of a few volts from a source of rather high internal impedance (see Philips Techn. Rev. **10**, 307-317, 1948).

1893: J. A. Lely and T. W. van Rijssel: X-Ray collimator producing a beam of very small divergence and large intensity (Acta crystallographica **2**, 337, 1949, No. 5).

Short description of an X-ray collimator, consisting of two polished glass bars, 120 mm long, placed almost parallel to each other (distance at one end 9 μ, at the other end 140 μ) and cemented between two glass plates. The narrow opening is placed as close as possible to the focus of the X-ray tube.

The primary beam entering through the slit is rendered parallel by multiple total reflections against the polished surfaces. In this way beams may be obtained with a divergence of only a few arc minutes, which suffices to measure lattice distances of 350-400 Å. With a slit the gain in intensity is 3-5 times and with a square aperture (pinhole) 10-30 times.

1894: E. J. W. Verwey: Nouveaux matériaux céramiques (Bull. Soc. Chim. France, Mars-Avril 1949, p. D 120-121). (New ceramic materials; in French.)

Survey of modern synthetic ceramic materials, which are used in electrotechnical engineering for their special dielectric, magnetic or conductive properties. Special attention is paid to the spinels containing iron (XFe_2O_4). See Philips Techn. Rev. **9**, 186-191, 239-248, 1947).

1895: E. J. W. Verwey: Valence induite (Bull. Soc. Chim France, Mars-Avril 1949, p. D 122). (Controlled valency; in French.)

By introducing monovalent ions (e.g. Li_2O) into a lattice of trivalent ions (e.g. NiO) a certain fraction of these ions passes into the trivalent state during the formation of the solid solution in equilibrium with an atmosphere containing oxygen. By this process the originally non-conducting oxide becomes a semiconductor, the conductivity of which can be easily controlled.

1896: E. J. W. Verwey: Arrangement des cations dans les spinelles (Bull. Soc. Chim. France, Mars-Avril 1949, p. D 123). (Cation arrangement in spinels; in French.)

The lattice energy of spinels is calculated as a function of small variations of the distance of the oxygen ions for different cases of occupation of the tetrahedral and octahedral interstices. Certain special cases are discussed in detail (see No. **1894** of these abstracts).

1897: W. Elenbaas and E. W. van Heuven: Water-cooled high-pressure mercury-discharge lamp for direct-current operation (J. Soc. Motion Picture Eng. **53**, 594-597, Nov. 1949).

A water-cooled high-pressure mercury-vapor lamp operated on direct current which has been used for motion picture projection is described. When mounted in a suitable reflector it is a powerful light source for high-speed photographic applications. The lamp has a bore of somewhat less than 2 mm and an arc length of $12^1/_2$ mm. It can be operated continuously at 1000 watts and has a brightness of about 320,000 candles per square inch along the axis of the arc. With the reflector described, an area 5 inches in diameter can be illuminated to 50,000 foot-candles with one lamp unit. With a lamp of double length consuming 2 kilowatts the illumination level may be increased considerably.

1898: P. C. van der Willigen: Lasbaarheid van zacht staal en laaggelegeerd staal (Lastechniek **15**, 317-322, 1949, No. 12). (Welding properties of mild steel and low alloy steel; in Dutch.)

From equilibrium diagrams of iron with different admixtures, such as Fe-S, Fe-P, Fe-H, Fe-N and Fe-O, the influence of these admixtures on the weldability of mild steel and low alloy steel is discussed; the part played by hydrogen from the electrode-coating is considered.

R 125: J. te Winkel: A note on the maximum feedback obtainable in an amplifier of the cathode-feedback type (Philips Res. Rep. **5**, 1-5, 1950 No. 1).

The total feedback on the last valve in a feedback amplifier is calculated for the case that the feedback voltage is derived from an impedance inserted between the cathode of the last valve and earth (cathode feedback). This quantity is shown to have an upper limit which depends only on the frequency and on the ratio of the transconductance and input capacitance of this last valve; the total feedback also appears to be invariably less than the product of the individual feedback of the valve and the feedback around the main loop.

R 126: J. L. H. Jonker: Valves with a ribbon-shaped electron beam: contact valve; switch valve; selector valve; counting valve (Philips Res. Rep. **5**, 6-22, 1950, No. 1).

This paper describes some of the results obtained in investigations with the object of developing cathode-ray tubes with ribbon-shaped electron beams suitable for purposes such as high-speed switching. By comparing the physical characteristics of the more common beams of circular cross-section with those of a ribbon-shaped beam, it is shown that by employing the latter the size of the tubes can be so much reduced that the customary radio-valve techniques can be applied in their construction. The new possibilities thus created are illustrated by (1) an electronic contact valve, which may serve as a telephonic switch, (2) a valve of similar design for replacing magnetic relays in central telephone stations, (3) a valve operating as a multi-contact switch, and (4) a valve capable of recording at a high speed a number of pulses fed to one of its electrodes.

R 127: J. M. Stevels: Some experiments and theories on the power factor of glasses as a function of their composition, I. (Philips Res. Rep. **5**, 23-36, 1950, No. 1).

For a number of series of glasses in which the composition is varied systematically, the power factor (tan δ) and the dielectric constant at 20 °C and for a frequency $f = 1.5 \times 10^6$ c/s have been measured.

R 128: J. D. Fast: The gas phase in equilibrium with a solution of carbon, oxygen, hydrogen and nitrogen in liquid iron (Philips Res. Rep. **5**, 37-45, 1950, No. 1).

From data available in the literature the composition of the gas phase in equilibrium with a dilute solution of C, O, H, and N in liquid iron is computed. In the range of temperatures from 1800 °K to 2000 °K the equilibrium pressures are approximated by formulae of the type

$$\log \frac{p_{A_nB_m}}{[\%A]^n[\%B]^m} = -\frac{C}{T} + D \quad . \quad . \quad . \quad (6)$$

where T is the absolute temperature and where the % figures indicate weight percentages of the various elements in the liquid. The form A_nB_m indicates the different gases (CO, CO_2, H_2, H_2O, CH_4 and N_2).

R 129: K. Rodenhuis: The limiting frequency of an oscillator triode (Philips Res. Rep. **5**, 46-77, 1950, No. 1).

This paper discusses the theoretical evaluation of the "limiting frequency", that is, the highest frequency at which a triode is capable of oscillating.

A triode is considered as a four-terminal network. General conditions are derived that must be satisfied by the four-pole coefficients in order that a four-terminal network shall be able to deliver power to an external circuit; these are then identified with the oscillatory condition of a triode. The relations existing between the four-pole coefficients of a triode on the one hand, and the properties of the electron current, the series resistances in the electrode leads, the resistance of the emissive coating, and the dielectric losses in the insulators on the other hand, are established and discussed. Finally the theory is applied to the ultra-short-wave oscillator triode EC 81 where the calculated limiting frequency is found to be 15% in excess of the limiting frequency observed. This discrepancy can be attributed to the approximations in the theory. The relative importance of various factors is clearly brought out by the calculations. In a concluding section the experimental determination of the limiting frequency is briefly described.

VOL. 12 No. 2, pp. 33-64 AUGUST 1950

Philips Technical Review

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

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

THE PHILIPS 100 kV ELECTRON MICROSCOPE

by A. C. van DORSTEN, H. NIEUWDORP and A. VERHOEFF.

621.385.833

There is hardly any other instrument receiving so much attention as that paid to the electron microscope by almost everyone practising the natural sciences: the physicist — creator of the science of electron-optics upon which this instrument is based — sees in the development of the analogy with the optical microscope a fine exemplification of the conception of the wave character of the material particles; the biologist, the analyst, the medical practitioner, the metallurgist and many others will be glad to have the use of this instrument; the engineer, finally, experiences in its practical execution an accumulation of heterogeneous technical problems which are (or may we say were?) a challenge to his ingenuity.

Fundamentals of the development

The step from optical microscopy to electron microscopy has brought mankind a great deal farther in the knowledge of the ultra-minute: in the extreme case the beam of light can only make objects of the order of about 0.2 micron (2000 ångström) distinguishable, whereas the electron beam has already made objects of 10 to 20 ångström visible, thereby permitting us to penetrate deeply into the fine structure of technically applied materials, into the world of bacteria and viruses, even into the field of large molecules, with but a relatively short and maybe not unbridgeable gap separating us from the individual atoms (of the order of 1 ångström).

Notwithstanding the successes already attained, electron-microscopy may be said to be still in its infancy. This is symptomized in the fact that the instrument with which one works still demands a great deal of the observer's attention, it being rather more an art than a science to make really good pictures with the electron microscope, pictures in which the very utmost is reached as regards resolving power, contrast, etc.

Defining briefly the fundamental idea of the electron microscope developed by Philips in recent years and now brought into production, we may say that the foremost aim in the construction of this instrument has been to facilitate work with the electron microscope as far as possible. The perfecting of the auxiliary apparatus and making it automatic, the systematic elimination of sources of possible errors, far-reaching simplification of all the manipulations necessary for operating the instrument, afford the investigator, on the one hand, the possibility of producing very good micrographs as routine work in a fraction of the time that was required to get a picture with former instruments, whilst on the other hand, if precautions are taken which require more time and skill, the acme of perfection can be reached.

The development of the Philips electron microscope has been based upon the fundamental work carried out by Le Poole and his co-workers in the Institute for Electron-microscopy at Delft. An experimental microscope, which was the result of that work, was described in this journal three years ago [1]).

Construction of the instrument

The instrument comprises four main parts:
1) the microscope tube, with, among others, the

[1]) J. B. Le Poole, A new electron microscope with continuously variable magnification, Philips Techn. Rev. 9, 33-45, 1947.

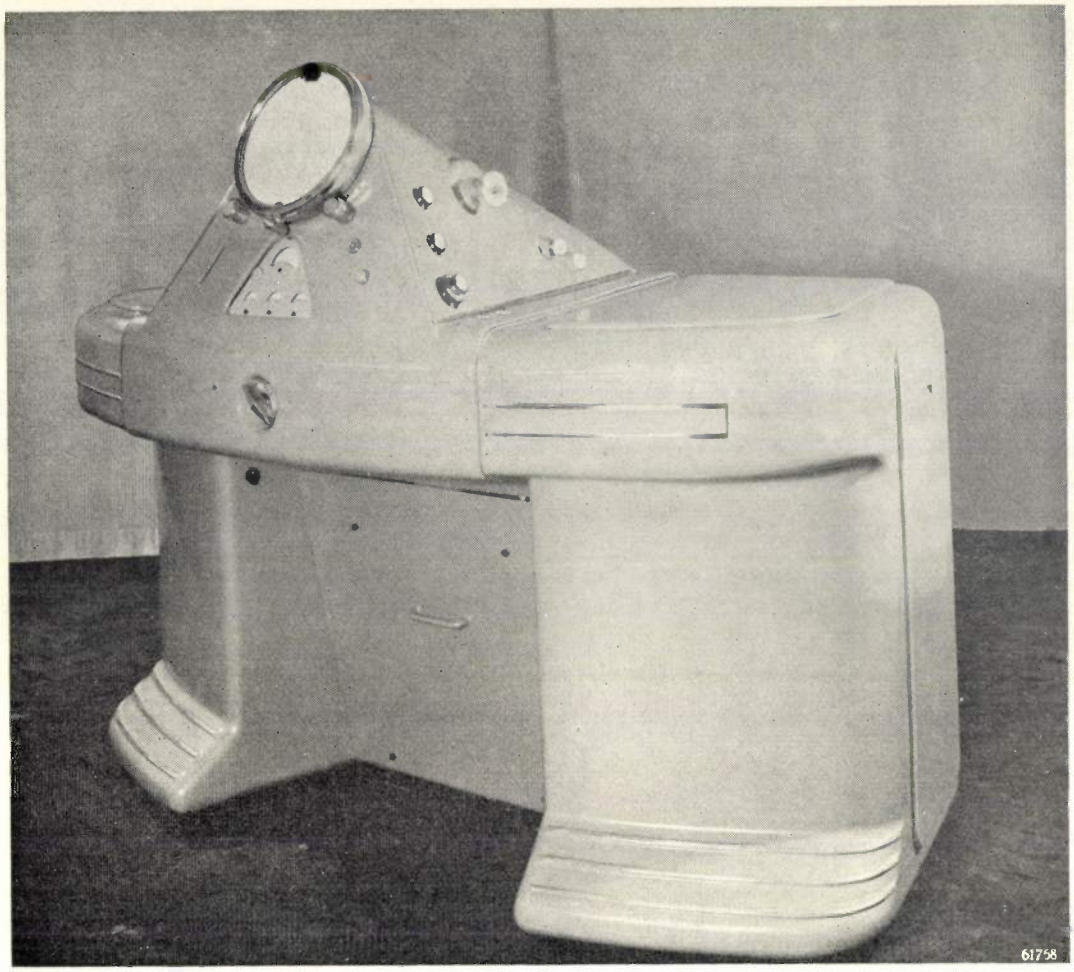


Fig. 1. The Philips 100 kV electron microscope (type number 11980). The microscope tube is mounted on a desk at an angle. The desk, comprising three compartments of cast aluminium, contains in the left-hand compartment the pumping system, in the right-hand compartment the supply apparatus for the lenses and all the other electronic apparatus, and in the middle, underneath the microscope tube, the high-tension generator, whilst on the panel underneath the circular viewing screen and on either side of the tube are various control knobs, control meters, etc.

electron source, the electron lenses — which are of the magnetic type —, the specimen holder, the viewing screen and the device for photographing;

2) a system of pumps for evacuating the microscope tube;
3) a high-tension generator supplying the voltage for accelerating the electrons;
4) the apparatus for energizing the electron lenses.

These parts, together with the necessary electrical gear, are built into a kind of desk shown in the photograph in *fig. 1* and in sectional view in *fig. 2*. The microscope tube is mounted at an angle on the desk with the uppermost end closed with a flat glass plate (16 mm thick in order to withstand the air pressure of 1 atmosphere when the tube is evacuated). On the inside of this glass plate is a fluorescent screen 20 cm in diameter, on which the enlarged image of the object is produced so that it can be observed through the glass.

In the following description of the various components most time will be devoted to the microscope tube itself, but also the principles of construction applied in the auxiliary apparatus deserve some attention. In conclusion details in the operation of the instrument will be passed in review, not in order to give directions for use but rather to show in how far the object outlined above has been attained.

The microscope tube

Principles of the electron-optical system

The formation of the image with an electron-optical system, in particular with magnetic electron lenses, has been discussed at length in the aforementioned article in this journal [1]), so that here

it suffices to recall the most important points.

A beam of electrons, emitted from a filament and accelerated in a strong electric field, is directed upon the specimen to be investigated. Different parts of the specimen attenuate the beam to a varying extent by scattering the electrons more or less. Thus the beam passing through it carries with it an image of the structure of the specimen, and after its subsequent passage through the field of an electron lens the plane of the specimen (object plane) is produced on a greatly magnified scale onto another plane (the image plane) situated farther along. The image can be made visible with a fluorescent screen in the image plane, but it can also first be magnified still further by a second and possibly still more electron lenses. The most commonly employed is the construction with two lenses (two-stage microscope), the first of which is called the objective lens and the second the projector lens.

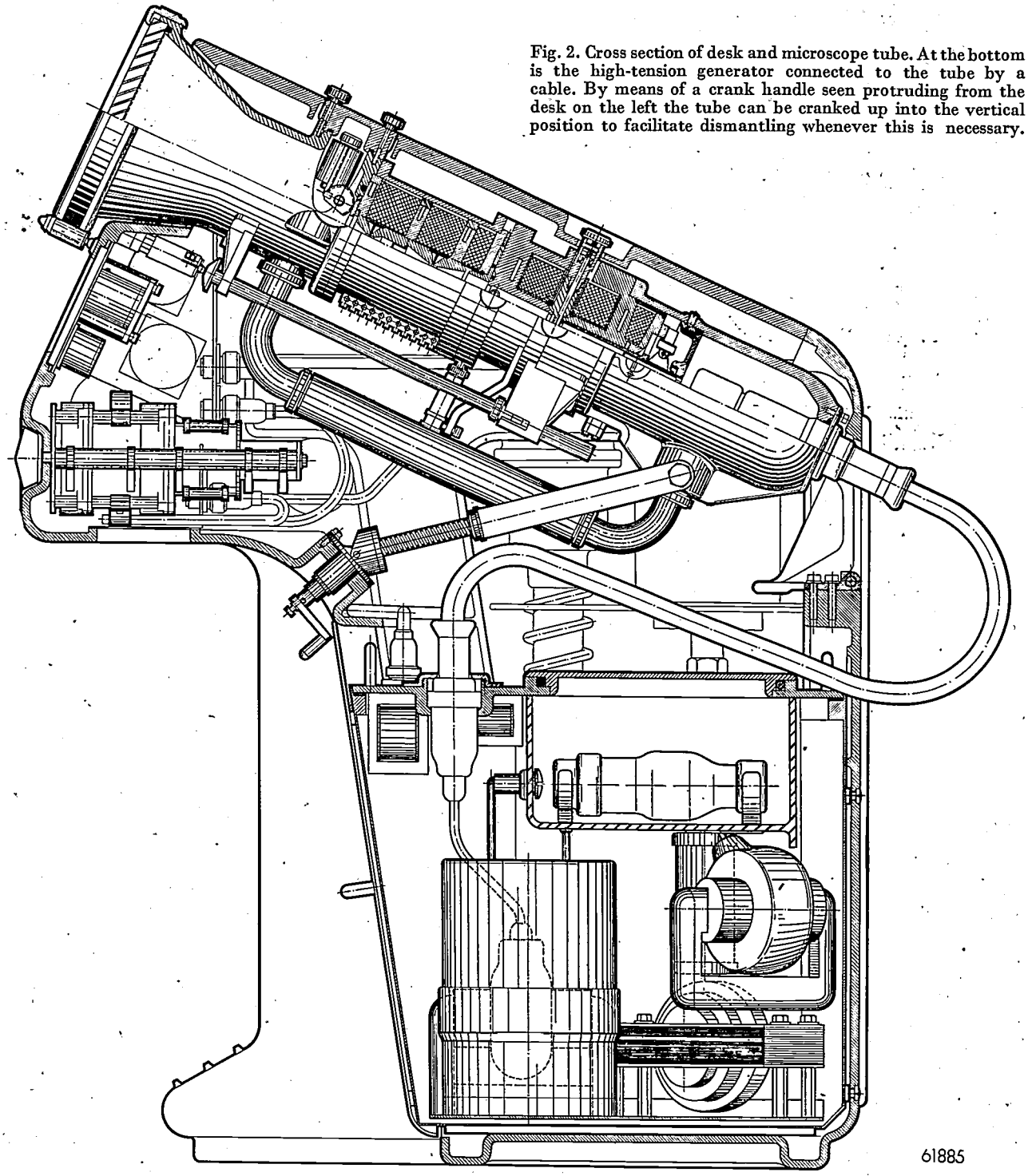


Fig. 2. Cross section of desk and microscope tube. At the bottom is the high-tension generator connected to the tube by a cable. By means of a crank handle seen protruding from the desk on the left the tube can be cranked up into the vertical position to facilitate dismantling whenever this is necessary.

A magnetic electron lens is formed by the inhomogeneous magnetic field in a short coil. For a strong lens (small focal length) strong fields are required, for instance of the order of 0.6 Wb/m² (6000 gauss). These are obtained by giving the coil a large number of ampere-turns and completely surrounding it by an iron jacket, except for an annular gap in the inner wall. To concentrate the field still more an annular pole piece of a suitable profile is inserted in the opening of the coil on either side of the gap; see *fig. 3*.

In this way focal lengths of 3 to 4 mm are obtained with a lens of reasonable dimensions and a magnification of the order of 100 times can be obtained. The focal length of the lens can be varied by varying the excitation.

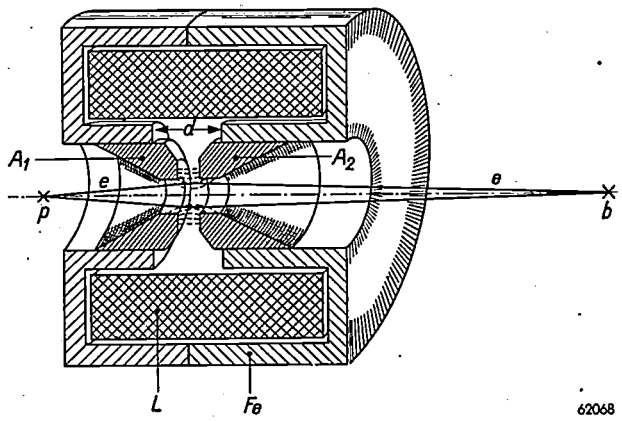


Fig. 3. Magnetic electron lens. *L* coil; *Fe* iron jacket with annular gap *d*; A_1, A_2 pole pieces; *e* electron beam. An image of an object at *p* is formed at *b*. Naturally the electrons have to travel in vacuum. To keep the space that has to be evacuated as small as possible, the annular gap is sealed vacuum-tight by filling it up with a non-magnetic material. — In actual fact the electrons do not travel along paths as drawn here: in the magnetic field they are moreover given a circling movement perpendicular to the axis, resulting in a helical orbit being followed. This is of importance when considering possible aberrations, but there is no need to consider it here; for this reason no account is taken of the circling movement, neither in this nor in the other illustrations.

The electron-optical system of the Philips electron microscope is mainly composed along the same lines as that of the experimental instrument designed by Le Poole. This system comprises not two but five magnetic lenses. The first of these acts as a condensing lens for "illuminating" the specimen, the second one is the objective lens and the last one the projector lens. The third and fourth lenses, which are energized alternatively, are denoted as the diffraction lens and the intermediate lens.

The intermediate lens makes it possible to reach a strong total magnification (in our case 60,000 ×) without the magnification of each of the three active lenses individually having to be excessively large; the main advantage is that the microscope is kept comparatively short (the overall length of the microscope tube is 81 cm). Moreover, by varying the excitation of the intermediate lens the magnification can be varied continuously within wide limits (from 4000 times to 60,000 times) with the whole area of the screen, which is 20 cm in diameter, still being completely covered. With a two-stage microscope it is hardly possible to get such a continuous variation of the magnification, as is explained at length in the article [1]) previously quoted.

By switching on the diffraction lens — a weak lens with wide bore — the enlargement is made smaller than that of the objective and projector together, and thus the range of enlargements from 1000 times to 4000 times is also covered, thereby linking up with the magnifications of the optical microscope. The diffraction lens derives its name from the fact that it affords a very simple means of obtaining an electron-diffraction diagram of the part of the specimen observed. This is achieved by adjusting the excitation of this lens in such a way that instead of the image plane of the objective lens its focal plane on the image side is displayed in the object plane of the projector lens (and thus onto the fluorescent screen). Just as in the case of the optical microscope, according to the theory already given by von Abbe, it is in this focal plane that the diffraction spectrum of the structure of the specimen appears. For the normal examination of a picture, in the path of the beam not far from the said focal plane, there is a very narrow diaphragm which allows only the first diffraction maxima belonging to the "coarse" structural details to pass through, which maxima have to contribute towards the formation of a well-contrasted image of these details. For obtaining diffractional patterns, however, it is the aim to obtain an image of the first diffraction maximum of each detail of the fine structure (arrangement of molecules and atoms). As these maxima lie much farther from the optical axis, the objective diaphragm just mentioned has to be replaced by a much wider one. Thanks to the presence of the diffraction lens, no other mechanical manipulations whatever are needed and it is therefore possible to change over in a few seconds from the normal image to the diffraction diagram of part of the specimen.

Fig. 4 gives a schematic representation of the paths followed by the electron rays in the three different situations: large magnifications, small magnifications and diffractional work. For the sake of clarity the width of the lens bores and of the electron beam is greatly exaggerated in these drawings: actually the path followed by the electrons

as far as the projector lens is nowhere more than about 1 mm from the axis, and it is not until they have passed through that lens that they diverge at large angles so as to cover the whole of the fluorescent screen.

nating current of 50 c/s, thereby causing the electron beam to "wobble", so that it still strikes the specimen in the same spot on the axis of the tube but at an angle varying in size and direction with a frequency of 50 c/s; see *fig. 5*. What this

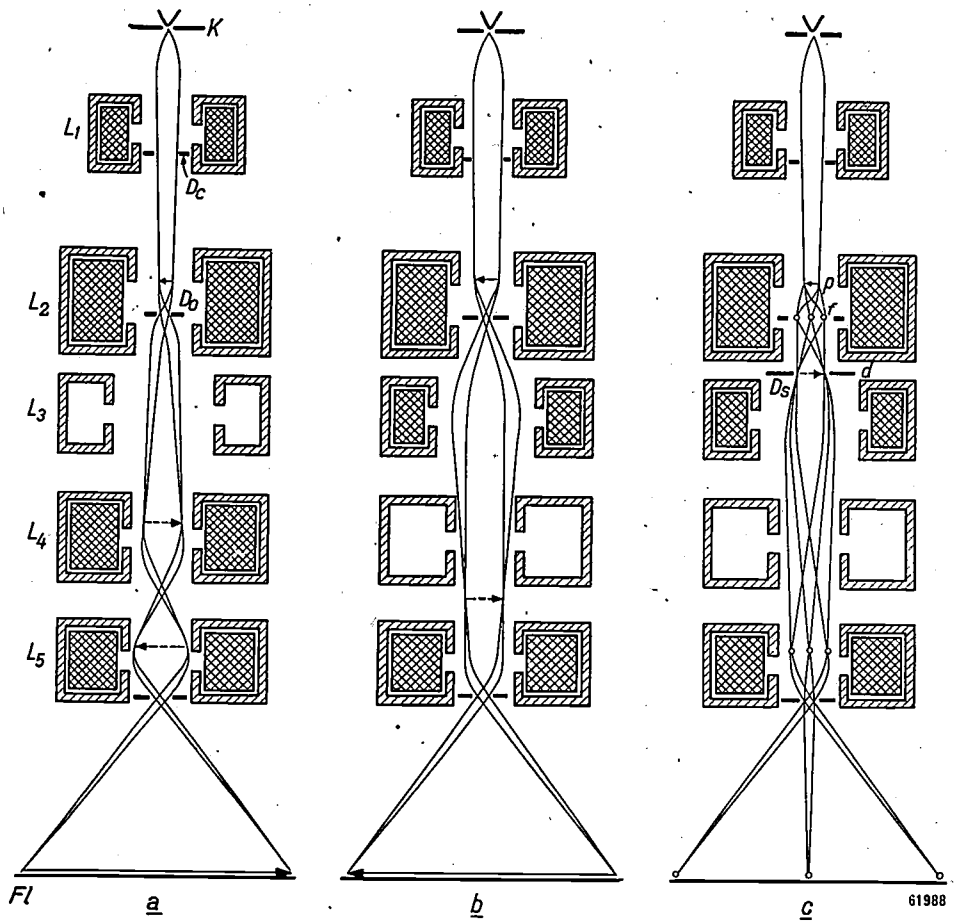


Fig. 4. Representation of the electron paths in the Philips electron microscope, which has five lenses: L_1 condenser, L_2 objective, L_3 diffraction lens, L_4 intermediate lens, L_5 projector; K the electron gun, F_1 fluorescent screen. The widths of the lens bores and of the electron beam have been drawn on a very much exaggerated scale for the sake of clarity. D_c, D_0 D_s are various diaphragms to be discussed later.
a) Situation with high magnification (4000-60,000 times); diffraction lens not energized.
b) With low magnification (1000-4000 times); intermediate lens not energized.
c) The situation for diffraction work. On the screen an image is produced of the focal plane f at the image side of the objective lens and not of the object plane p. An intermediate image of p is formed in the plane d.

In addition to the principle of the intermediate lens and the diffraction lens there are two other important developments used in our electron microscope which are due to Le Poole and which will be described here, namely the focusing device and the manner in which the image is photographed.

The focusing device consists, in our case, of two sets of deflection coils mounted between the condenser lens and the specimen. While observations are being made these coils are out of action, but for focusing they are energized by an alter-

amounts to is that a single-plane beam is temporarily used with a much larger aperture than is usual (1/100th radian as against the normal 1/1000th radian). The blurring of the image caused by a small deviation between the specimen and the object plane of the objective lens is made much more pronounced by the larger aperture; when the excitation of the objective lens is varied in order to bring the object plane closer to the specimen the minimum of the blurring thus gives a distinct criterion for the coincidence of the object plane

with the specimen. Focusing with the "wobbling" beam is of particular advantage for the largest magnifications, when the images on the fluorescent screen have rather low luminosity.

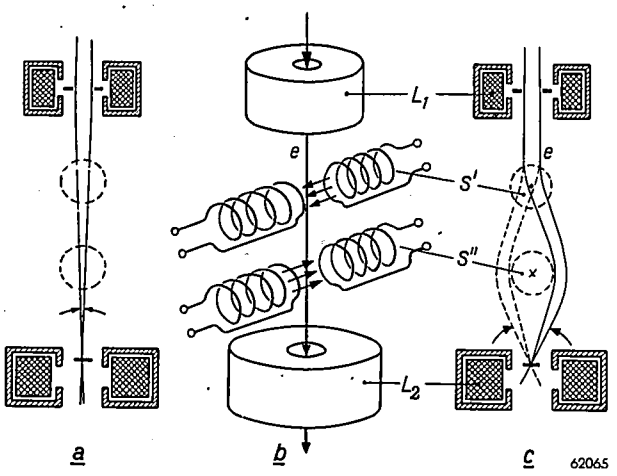


Fig. 5. Schematic representation of the focusing system. Between the condenser L_1 and the objective L_2 are two small coils S', S'', generating two magnetic fields in opposite directions and at right angles to the electron beam e (b). Normally these coils are not energized; the beam passes straight through and has a very small angle of incidence (a). Upon the coils being energized with a direct current the beam e is deflected by S' and brought back again by S'', in such a way that it strikes the specimen at the same point as before (c, the fully drawn lines). While focusing the coils are fed with an alternating current of 50 c/s; a "wobbling" beam is thus obtained (dotted lines in the drawing), which is equivalent to "illuminating" the specimen with a beam having a much larger angle of incidence than normal.

For photographing an image the electron beam is made to fall directly upon the photographic material. Following Le Poole's method, a roll-film is used which is placed between the screen and the projector lens, fairly close to the latter. The photographs thereby obtained are much smaller than the image on the screen, thus effecting a saving in the cost of the film and in the space occupied in archives, etc., whilst at the same time, owing to the greater concentration of the available electron energy, a relatively short exposure suffices. If necessary the photographs can subsequently be examined under a large optical magnification; thanks to the fine grain of the film nothing is lost of the resolving power of the electron microscope.

This method is made possible by the minuteness of the relative aperture at which the projector lens is used: the individual beams each producing an image point on the screen are so narrow that an astonishingly great depth of focus is obtained. The image is sharp from a few cm behind the centre of the projector lens right up to the screen and would in fact still be so at several metres beyond the screen.

The electron gun

The electrons are supplied by a V-shaped filament made of tungsten, surrounded by a metal cap with an opening of about 1 mm diameter through which the electrons can emerge; see *fig. 6*. This assembly is at a high negative potential (40 to 100 kV) with respect to the earthed metal shield of the microscope tube. The filament is fed by a transformer connected to the mains and insulated against 100 kV. The electrons emerging through the opening in the cathode cap are accelerated by the prevailing potential difference and, under the focusing action of the local electric field, follow in a narrow beam the axis of the tube.

The field required for focusing is obtained by giving a suitable shape to the cathode cap and the shield and keeping the cap at a negative potential of 80 to 100 V with respect to the filament. The latter is ensured by connecting the cap direct to the high-tension point a resistor being inserted between that point and the filament. At the same time this provides for a self-regulating action of the emitted electron

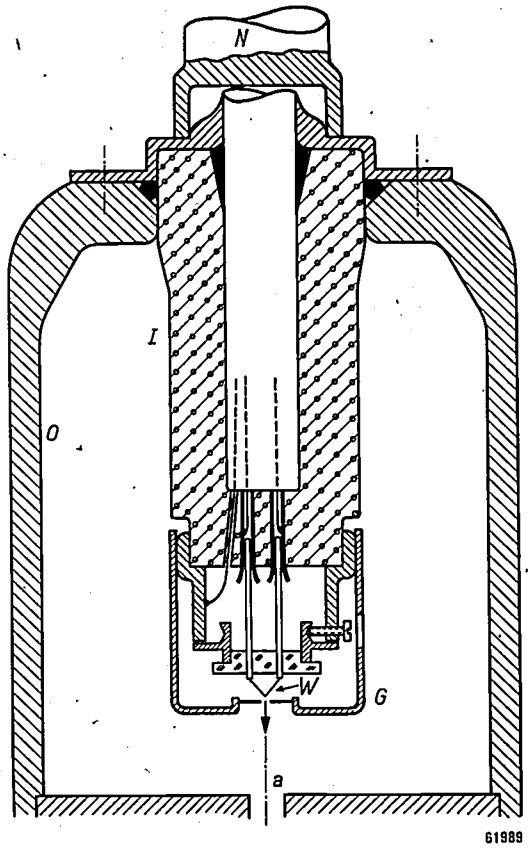


Fig. 6. Construction of the electron gun (simplified). W filament, G metal cap with small opening through which the electrons emerge. Filament and cap are carried by the cylindrical insulator I made of a special kind of "Philite", mounted vacuum-tight in the head of the earthed metal shield O of the microscope tube. N high-tension cable with rubber insulation and earthed shield, a entrance of the miscroscope tube proper.

current: as the emission current increases, the negative potential of the cap (acting as control grid) with respect to the filament increases, thus counteracting the increase of the current (according to the principle of a self-biased triode). We shall presently see that this regulating action is of great importance for the construction of the high-tension generator. The emission current with which one normally works is from 10 to 20 μA, of which only a very small fraction, say 10^{-10} A or still less, reaches the fluorescent screen at the other end of the microscope tube.

Fig. 7. Replacement of the filament. The removed cathode cap is seen standing on the desk above the man's hand.

Owing to the local relatively high temperature, when the electron microscope is in normal use, the filament has to be renewed every few days. This is very easily done, since the filament is mounted on a sintered glass base with fused-in plug pins which can easily be inserted in or taken out of the holder. The filament is reached by turning down a hinged cover protecting this end of the microscope tube, unscrewing the head of the shied and drawing the electron gun out (*fig. 7*). Each filament is previously centred on its base, so that after it has been properly plugged in the point of the V is automatically aligned in the opening of the cathode cap. If necessary, after the cap has been put back, the position of the point of the V can be corrected by slightly shifting the filament holder with the aid of three set screws.

The lenses, diaphragms and apertures

The electron beam originating from the electron gun enters the tube of the electron-optical system, first passing through the condenser. This relatively weak lens concentrates the beam more or less upon the specimen. The focal length can be adjusted between 30 and 2.5 cm. The angle of incidence of the electron beam striking the specimen varies between 0.002 and 0.00015 radian. A diaphragm in the condenser, 0.3 mm wide, limits the width of the beam. When the cathode end of the microscope tube is opened (fig. 7) this condenser diaphragm can easily be withdrawn from the condenser with a special tool, for cleaning it when necessary.

The next lens is the objective lens, a strong lens with a focal length of about 4.5 mm. In the conventional construction this lens has two pole pieces approaching each other to within 1 to 2 mm and having an axial bore of 1 to 3 mm through which the electron beam passes. In our microscope, however, the distance between the pole pieces and also the width of the bore is approximately 10 mm and the pole pieces have been made correspondingly heavier. The whole lens is "scaled up" as it were, as may be seen from *figs 8a* and *b*. This method of construction of the objective lens has three important advantages. In the first place the spherical aberration, one of the factors limiting the resolving power of the microscope, is appreciably reduced as compared with the normal construction. Secondly another important aberration, astigmatism, can be practically eliminated owing to the fact that the 1 cm bore can be mechanically worked with smaller relative deviations from the rotational symmetry

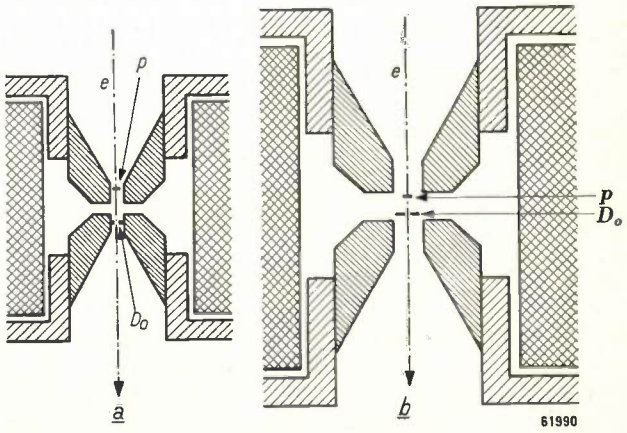


Fig. 8. *a*) The conventional construction of the objective lens. *b*) "scaled up" objective lens; *e* electron beam, *p* specimen, practically in the object side focus D_0 objective aperture.

than is possible with a bore of 1 to 2 mm, whilst magnetic inhomogeneities in the material of the pole pieces are almost entirely eliminated as a source of asymmetry thanks to the greater distance between the iron and the electron beam. Thirdly — and perhaps this is the most important advantage owing to the constructional consequences — the focal point of the object, which in the normal construction is always somewhere inside the bore of the upper pole piece, now lies in the space between the two pole pieces (fig. 8*b*). Thus the specimen, which has to be brought into a plane very close to the focal point of the object, can be brought into place by means of a straight rod from the side, whereas in the conventional construction a rather complicated mechanism was always needed to bring the specimen over the objective lens in the tube and then to lower it into position in the bore. Not only is the specimen holder thus made much simpler in our construction, but it is now also a practical possibility to make a very efficient "specimen lock", thereby reducing the time taken in exchanging the specimen and re-evacuating the microscope tube to only about 20 seconds. The construction of the specimen holder will be further described later [2]).

The objective aperture is likewise situated in the space between the pole pieces of the objective lens (fig. 8*b*). As already mentioned, two different apertures can be used, a small one for normal imaging (40μ) or a larger one for diffraction work (about 1 mm). Both apertures are contained in a platinum strip fixed to a rod-like holder which is also inserted from the side. It depends upon the depth to which the holder is screwed in, which of the two apertures is brought into the beam.

The position of the apertures can very easily be checked by switching on the diffraction lens of the microscope, when an image of the small aperture can be seen on the screen magnified about 200 times. Also any contamination of the diaphragm due to particles broken away from the specimen is immediately detected. Should the aperture need cleaning then the holder can be unscrewed and drawn out, without involving any dismantling of the microscope.

[2]) A "scaled-up" objective lens was already used by L. Marton (Phys. Rev. 58, 57, 1940). From theoretical considerations W. Glaser (Z. Physik 117, 285, 1941) has drawn the conclusion that in the normal construction of an objective lens the object plane can come to lie between the pole pieces when the lens is very strongly excited, and that the spherical aberration is then comparatively small. In the normal construction, however, the difficulty is encountered that with such a strong excitation the pole pieces become saturated. This does not occur in the "scaled up" construction with the excitation required here.

Contrary to the conventional construction, where the objective aperture is built in, inaccessible from the outside, and cannot be directly checked, with our microscope there need never be any uncertainty as to the condition and centering of the aperture, and as far as this is concerned one can always work under optimum conditions.

Between the objective lens and the next one, the diffraction lens, is a square selector diaphragm the size of which can be varied continuously. This is used for diffraction work, for selecting the part of the specimen of which it is desired to obtain a diffraction pattern. In the plane of this diaphragm an intermediate image of the specimen is then formed; upon the diaphragm being reduced the electron rays emerging from the outer parts of the specimen are prevented from taking part in the formation of the diffraction pattern. With the smallest opening of the diaphragm that can be used the effective plane of the specimen measures 1μ × 1μ and with the largest opening about 30μ × 30μ. The principle of the construction of the selector diaphragm is represented in *fig. 9*.

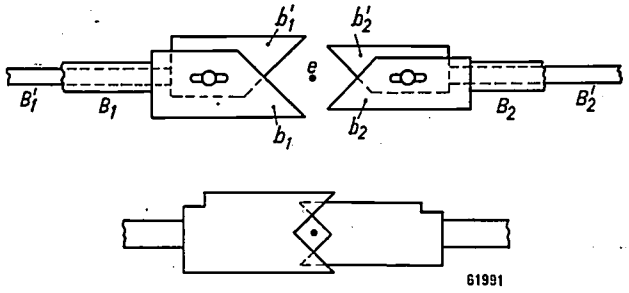


Fig. 9. Schematic representation of the construction of the selector diaphragm. e = axis of the microscope, perpendicular to the plane of drawing. Two fork-shaped pieces can be made to overlap more or less by means of the rods B_1, B_2. The square opening in between the forks is thereby varied continously in size and can be shifted in its entirety tothe right or left. The opening can also be displaced upward or downward (in the plane of drawing) by changing the position of one of the two wedge-shaped plates (b_1', b_2') of each fork with respect to the other (b_1, b_2) in the direction of the rods, by means of two other thin rods B_1', B_2' sliding in an axial bore of the first-mentioned rods. All four adjustments are made from the outside without disturbing the vacuum in the microscope tube; the adjustment is checked by projecting an image of the square on the fluorescent screen.

After passing the selector diaphragm the electron beam passes through the diffraction lens and the narrow bores of the intermediate and projector lenses. The functions of these various lenses have already been described in the foregoing, and their construction does not present any new aspects. All five lenses are provided with water-cooling for carrying off the heat generated in them by the excitation currents. To illustrate this

description a (somewhat simplified) cross-section of the whole of the microscope tube is given in *fig. 10*.

It is a problem of great practical importance how to ensure (*a*) that the electron beam passes freely through the numerous narrow openings and (*b*) that moreover the beam follows exactly the axis of each lens in order to avoid the aberrations inherent in rays oblique to the axis (coma, astigmatism). This latter point is of particular importance for the objective lens. The two requirements together imply in the first place that the axes of all five lenses must be accurately aligned. In our microscope this has been ensured in the following way. The condenser lens and the two halves of the objective lens — this has to be made in two halves to allow of the specimen table (dealt with later) being put into place — are joined together with precisely gauged fittings. The coils of the diffraction lens, intermediate lens and projector lens are wound on one common iron-core cylinder with magnetically insulating rings in between, the pole pieces for all three lenses, together with the necessary spacing rings, being inserted in the bore of the cylinder (fig. 10). In this way these last three lenses form one rigid unit, which in turn is connected to the first-mentioned part of the tube with a fitting, whilst the first pole piece of the diffraction lens, forming the link between the two parts of the tube, can be readjusted with set screws if necessary after the instrument has been assembled. The axes of the five lenses are thus so accurately aligned that any further adjustment on the part of the user of the microscope could not improve matters, so that there was no need to make any provision for this possibility. In the second place, however, it has now to be ensured that the electron beam entering the tube is directed precisely along the axis of the tube. Since the electron source has to be repeatedly renewed, it must be made possible for the user of the microscope to centre the beam anew when necessary.

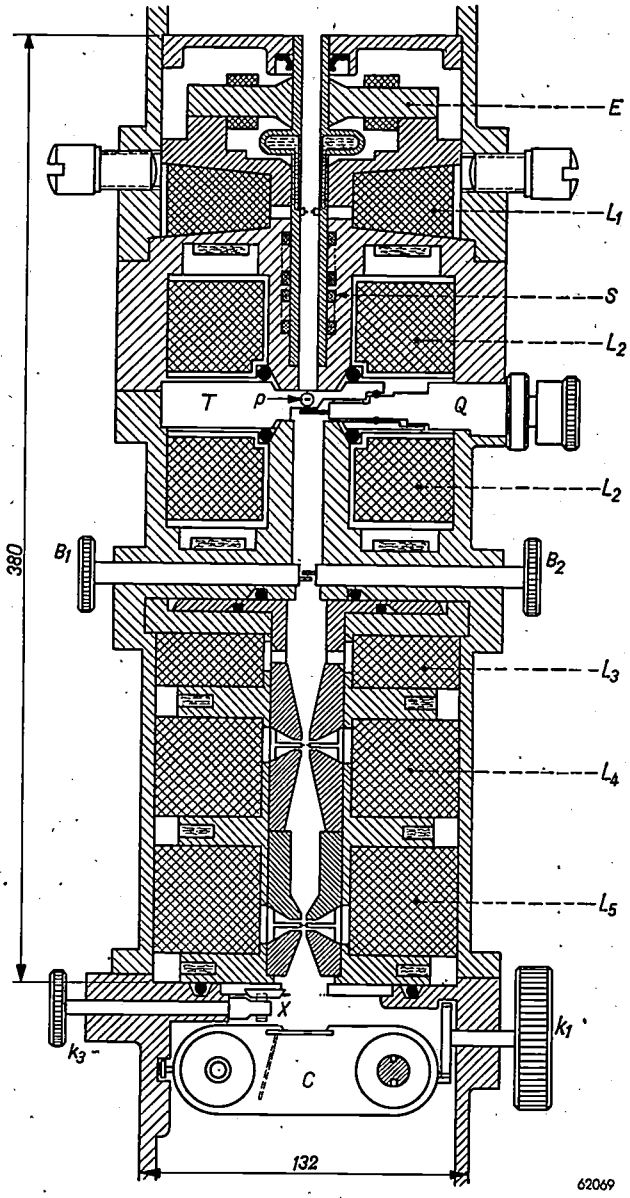


Fig. 10. Cross-section (simplified) of the microscope tube with the electron-optical system consisting of five lenses. L_1 condenser, S focusing device, L_2 objective lens (made in two parts), Q objective-aperture holder. This holder, and the specimen holder p, which cannot be seen in this drawing because it stands at right angles to the plane of the cross section, are illustrated in detail in fig. 13. B_1, B_2 holders of the selector diaphragm, L_3 diffraction lens, L_4 intermediate lens, L_5 projector lens. E centering device. (T specimen table, C camera with turning knob k_1, k_3 knob for the marking device X; see in the following text.)

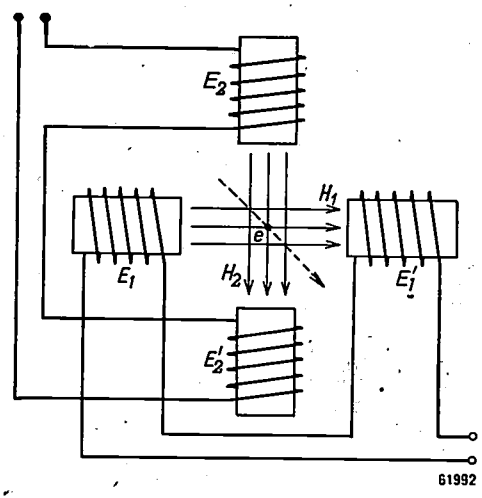


Fig. 11. Schematic representation of the centering device. e = electron beam perpendicular to the plane of drawing. The coils E_1, E_1' connected in series generate a magnetic field H_1, the coils E_2, E_2' a field H_2. The electron beam is deflected by the Lorentz force in a direction perpendicular to the resulting field at the place of the beam. By slightly varying the current through the two pairs of coils the resulting field can be varied in strength and direction and thus the electron beam deflected as far as may be necessary in any desired direction.

As a departure from other constructions, where the electron gun can be shifted or tilted in its entirety by means of flexible connections between the electron gun and the microscope tube, we have chosen the constructionally more attractive method of directing the beam by deflection with the aid of magnetic fields. This centering device can be seen at the top of the diagram in fig. 10 (*E*). It consists, in essence, of two pairs of coils, each pair supplying a magnetic field directed transversely to the beam. Its action is explained in *fig. 11*.

The specimen holder

The specimen is mounted on a very small silver plate with a rectangular opening and this carrier is clamped in the rod-like holder already mentioned (*fig. 12*). The holder is placed in what may be called the specimen table, a flat metal disc of non-magnetic material placed in the gap between the pole pieces of the objective lens. This disc can be shifted about 1 mm, by means of two pins, in two directions at right angles to each other and both perpendicular to the axis of the microscope tube. By a combination of these two movements any part of the small specimen can be brought into the centre of the object plane, thus making it possible to "scan" the specimen with a uniform movement. Rubber ring seals over and underneath the disc ensure that the vacuum is maintained. This sealing and some other constructional details of the specimen table and holder (and also of the holder for the objective aperture already discussed) are given in *fig. 13*.

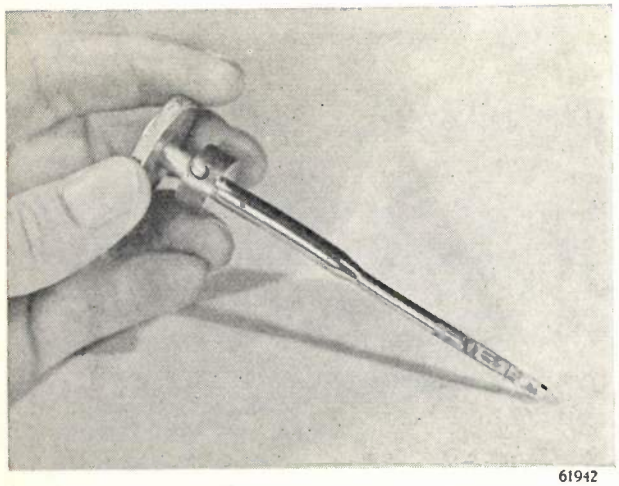


Fig. 12. The rod-like specimen holder, with underneath on the left (greatly magnified) one of the small silver plates acting as carriers for specimens. The drawing on the right below shows how such a plate is clamped in the holder.

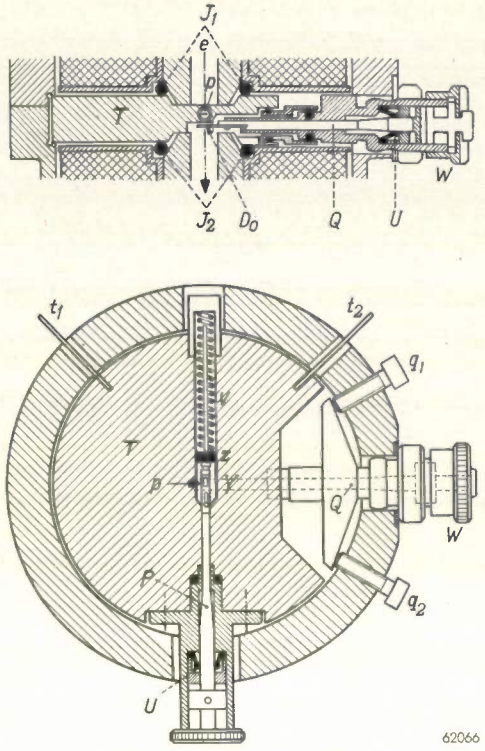


Fig. 13. The specimen table (*T*), into which the rod-like specimen holder *P* is inserted. t_1, t_2 pins for shifting the table, the vacuum being preserved by the rubber ring seals J_1, J_2. *Y* specimen lock with rubber piston *z* and spring *v*. *Q* holder of the objective aperture D_0, which is only a few mm away from the specimen. The axial displacement of this latter holder, for changing the two objective apertures, is brought about by turning the milled nut *W*, two stops being provided, each corresponding to the exact position of one of the apertures. In order to facilitate accurate centering of the small aperture in the axis of the microscope the guide for the holder rod *Q* can be swung round a little; this adjustment of the direction of *Q* is made with the set screws q_1, q_2. It is only needed if the holder has been entirely withdrawn, for instance for cleaning the diaphragm. *U* are rubber rings of the so-called oil-seal type.

The two pins just mentioned are operated with two knobs at the side of the fluorescent screen via a system of levers either side of the microscope tube. Thus the specimen table can be shifted while one is sitting in front of the desk and watching the image on the screen. The displacement of each pin can be read accurately to within 1μ on an illuminated dial, so that it is quite easy to find a certain part of the specimen again. The lever system is so constructed that in the event of any slight deformations of the desk (displacements of the lever bearings) no force is transmitted to the pins; thus the image is not shifted if, for example, someone should lean upon the desk! The specimen is shifted quickly and very smoothy, the frictional force to be overcome in moving the specimen table being no more than about 10 newton (1 kg).

It has already been mentioned that the specimen holder has a sort of "lock". The principle of this is to be seen in fig. 13. When the specimen holder is withdrawn a rubber piston is pressed by a spring into the opening through which the specimen holder is drawn out, thereby automatically closing that opening. When the holder is reinserted the piston is forced back, against the pressure of the spring. Only the small amount of air between the specimen rod and the bore in the table (some mm^3) gets into the microscope tube.

The rod-shaped specimen holder has also made it possible to devise an extremely simple method for making stereo-micrographs. Without disturbing the vacuum the rod can be turned over a small angle about its axis. This axis lies in the specimen plane, so that, as the rod is turned, the centre of the specimen remains in place. In this way two micrographs can be taken of the specimen in succession, differing only in that the electron beam passes through the specimen volume at a somewhat different angle. When these photographs are viewed in a stereoscope one sees a three-dimensional picture. Thanks to the great depth of focus of the electron microscope, corresponding parts in the two photographs are equally sharp, although the concerned part of the specimen was once in front of and once behind the exact object plane.

The camera and accessories

It is a remarkable fact that in electron-microscopy the photographically recorded picture shows more than the visual image on the fluorescent screen, such in contrast to other comparable instruments like the telescope and the optical microscope, where usually the reverse is the case. It is therefore only obvious that where an electron microscope is used one prefers the photographic method for purposes of observation. This also has the advantage that the instrument need not be kept engaged for studying the pictures, whilst moreover the pictures recorded in micrograms can always be referred to again.

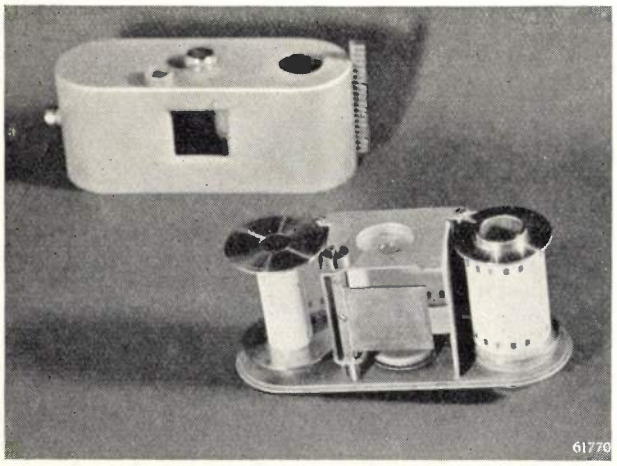


Fig. 14. The camera for 35 mm film used for recording the image.

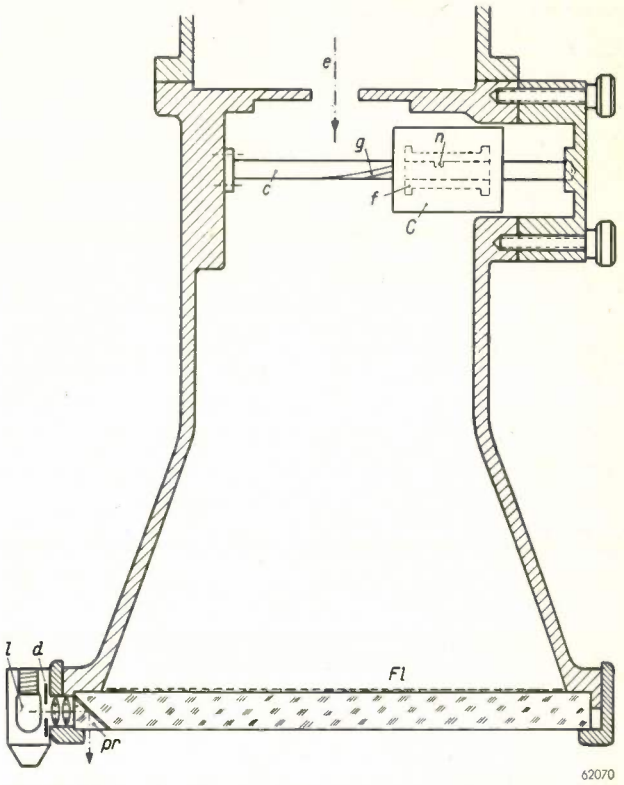


Fig. 15. Projection chamber of the microscope. Here the camere *C* is shown in the position of rest. To make a micrograph the camera is brought into the electron beam *e* by sliding it along the pin *c*. At the bottom of the drawing, on the edge of ths glass cover is a simple photometer for roughly determining the exposure time required for each picture. Via a small lens and a prism *pr* contained in a corner ground out of the glass cover, a small lamp *l* gives a spot of light seen by the observer beside the image on the fluorescent screen (*Fl*). The strength of this spot of light is varied with a diaphragm *d*, adjustable in four steps, until it corresponds as closely as possible to the brightness of the image on the screen. Once the sensitivity of the film is known the position of the diaphragm — the figures are illuminated by the lamp itself, so that they can be read in the dark — gives a measure for the exposure time required, the four positions corresponding to exposures in the ratio of 1 : 3 : 10 : 30. In practically all cases a more accurate determination is not needed. (For *f*, *g* and *n* see fig. 16.)

In the case of our microscope the images are photographed on standard film of 35 mm width. The camera, illustrated in *fig. 14*, contains the two spools for winding and unwinding the film and a shutter in the form of a hinged cover. The film comes to lie at such a distance from the projector lens that the photographed image corresponds to a square of 14 × 14 cm on the screen, i.e. the square just fitting in the circle of 20 cm diameter. The spools offer space for a roll of film sufficient for 40 micrographs.

While visual observations are being made the camera is turned aside from the projection chamber, but still within the vacuum; see *fig. 15*. When a micrograph has to be made the camera is moved along into the electron beam by means of a knob provided for the purpose outside the microscope tube. With a second knob the shutter is opened and the electrons are thus allowed to fall upon the film. The exposure required is roughly determined by a simple, small, measuring device shown at the bottom of fig. 15.

The film is transported automatically with the movement of the camera: every time the camera is moved forward and back again the film is wound one picture frame farther, so that double exposures are precluded. The mechanism employed for this is similar to that used in a wimble, a to-and-fro movement being translated into a unilateral rotational movement. The mechanism is illustrated and further explained in *fig. 16*. When the film is completely exposed air is let into the microscope and the camera taken out, after removing a cover held with two screws.

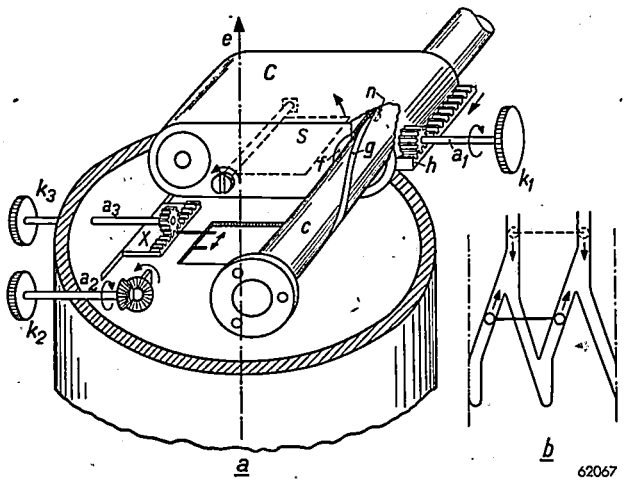


Fig. 16. Mechanism for operating the camera and automatic film transport (here the electron beam e is directed upward). *a*) The camera C is moved to and fro by means of the knob k_1 via the vacuum-sealed spindle a_1 and the ratchet-gear drive h. The knob k_2, which is turned with the other hand, opens the shutter flap s, via the spindle a_2. When the camera is moved it slides along a fixed pin c passing through the hollow bush of the film spool f on which the exposed film is wound. On this bush is a cam n directed inward which engages in a coiled groove g cut in the fixed pin. As the camera is moved along the pin the spool is thus caused to rotate, such that the film is wound up exactly half a picture length farther while the camera is travelling from the outermost position as far as it can go inwards. Upon the camera being returned to the outermost position again the cam follows a coiled groove in the opposite direction (leading back to the beginning of the first groove), so that the spool is rotated further in the same direction and the film is wound up another half picture length. The cam can never return along the same groove because of a freewheel coupling preventing a corresponding reverse rotation of the winding spool; one is therefore obliged to move the camera right along to the stop in both directions, thus precluding any possibility of partial overlapping of the photographs (double exposure). A counting mechanism on the knob k_1 shows how many of the 40 picture frames of the film have been exposed.

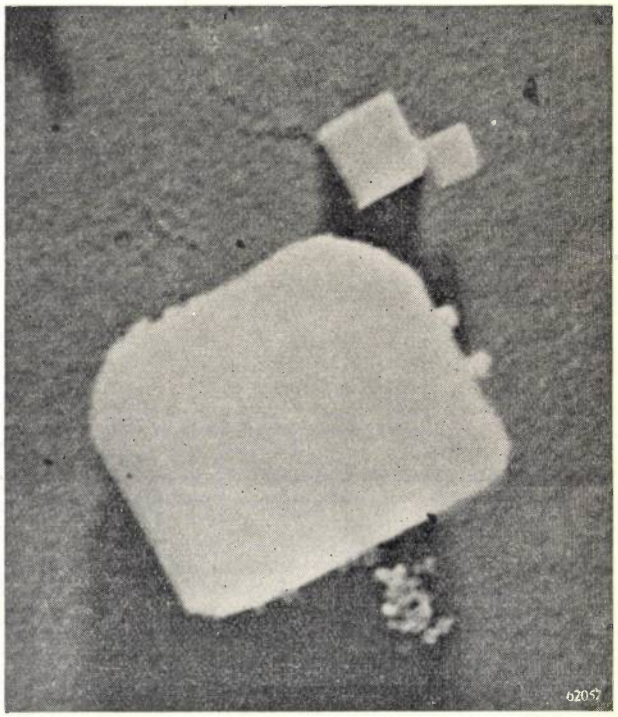
As a matter of fact there are two cams, and not one, running in two grooves in the one direction and in two opposed grooves in the other direction (see development of the pin in fig. 16*b*). Thanks to this double cam action only half a revolution of the film spool corresponds to the transport of a whole picture length; this spool therefore has twice the diameter, which has the advantage that there is relatively little variation in the length of film wound up from the first to the fortieth picture and thus little film need be wasted. Moreover the steeper grooves give a greater lateral force component on the cam, making it easier for the film spool to be rotated. X is a device for marking the scale on the micrographs.

Of course, if it is desired to develop and examine a certain micrograph at once, the camera can be taken out before the film is completely exposed. While the work in the dark room is being done one can carry on with the microscope work with another camera previously prepared.

The scale can very easily be marked on each image that is to be photographed. Immediately underneath the projector lens, in the electron beam, are two parallel wires, the distance between which can be changed by means of a ratchet-gear drive (see X in fig. 16, with knob k_3 and spindle a_3). These wires cast two parallel shadow marks on a cm scale drawn close to the edge on the fluorescent screen. The distance between the wires is adjusted so that the distance marked off on the cm scale corresponds to 1 μ length in the specimen. The distance required between the marks follows directly from the magnification factor shown by a milliammeter measuring the excitation current of the intermediate lens or the diffraction lens. The 1 μ length thus marked off is, of course, correct at any point where the electron beam is intercepted, thus also on the film. In *fig. 17* a micrograph is reproduced which was made with our electron microscope and which clearly shows the 1μ marking.

For good micrographs to be obtained a very high degree of electrical and mechanical stability of the microscope is essential. The electrical stability will be discussed presently. As regards mechanical stability we have to stress once more the importance of the rigid construction of the microscope tube, in which no flexible connections of any kind have been used between the various parts. Even when all possible precautions have been taken (some of these will be mentioned) the transmission of small vibrations to the microscope tube cannot be entirely avoided. In order to test the sensitivity of our instrument to vibrations we introduced a vibrator in the desk by way of experiment, when it was found that the specimen holder vibrated with an amplitude of 2 μ. Although, with a magnification of 20,000 times, this should have resulted in an unsharpness of 4 cm on a fixed film, the micrograph obtained was still perfectly sharp! Owing to the rigid construction the transmitted vibrations apparently caused all parts of the microscope tube, including the specimen holder, to vibrate as one whole, thus without affecting the picture quality at all.

a



b

Fig. 17. Micrograph of magnesium oxide taken with the Philips electron microscope at a magnification of 18,000 times. The cube-shaped crystals lie on a film of "Formvar" and the whole is "shadowed" with a stream of uranium vapour. At the top on the left-hand side of the photograph are the two scale markings, the distance between which represents a length of 1 μ. Photo *b* shows a detail of *a* at a magnification of 60,000 times.

A closer examination of the original micrograms shows that here the resolving power is better than 50 ångström. With factory-made microscopes built according to the design described here micrographs have already been obtained where the resolving power was found to be 25 ångström.

In conclusion of this part of the description of the electron microscope, in *fig. 18* a photograph is given in which the whole of the microscope tube is seen and in which some of the parts discussed can easily be identified.

The high-tension supply

The electrons can be accelerated with a voltage of 40, 60, 80 or 100 kV. The highest voltage — which is unusually high for an electron microscope of a commercial type — makes it possible to photograph even the thicker objects (viz. bacteria) occurring in normal investigations. The lower voltages may be required for examining extremely thin and light specimens, which at a higher voltage give pictures of less contrast. The lowest voltage may also prove useful for making micrographs under unfavourable vacuum conditions, for instance when the specimen or the film is giving off gas, which at a higher voltage would make it impossible to operate the tube.

In this connection it is well to touch upon a fundamental question, namely that of the brightness of the magnified images produced on the fluorescent screen. In the case of micro-projection with an optical microscope with the maximum useful enlargement of about 1000 times, the pictures are still sufficiently bright if the light source used has a brightness

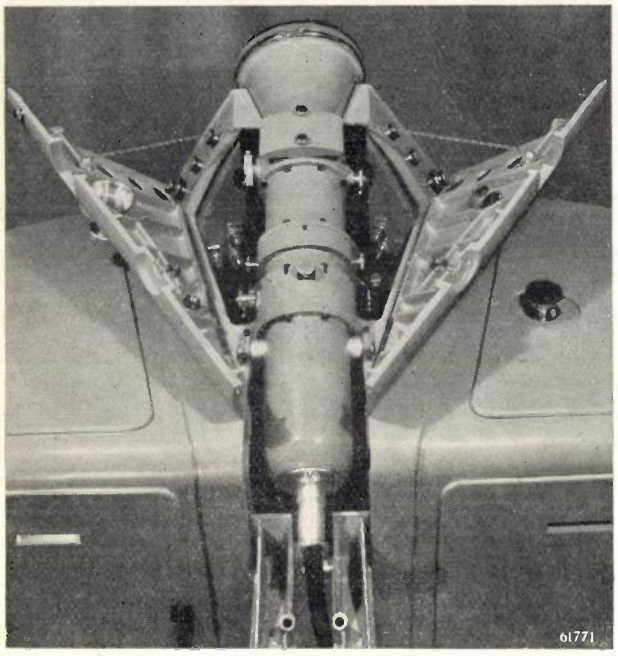


Fig. 18. Top view of the desk with the protective covers of the microscope tube thrown open. At the top the conical projection chamber, in the neck of which are the two knobs for operating the camera and the removable cover for taking the camera out. Halfway down the tube, on top, is the knob for the specimen holder, to the left of that, at the side of the tube, the knob for the objective aperture, and a little higher up, on either side of the tube, the knobs for the selector diaphragm.

of say 2000 candela/cm². For a 100 times larger magnification, however, a 10,000 times brighter light source would be required to get pictures of the same brilliancy, thus a light source about 100 times brighter than the sun! In electron-microscopy the situation is essentially analogous. To get a sufficiently bright fluorescent image with a magnification of 100,000 times an enormously "bright" beam is needed ("brightness" here is the energy per second, per unit of solid angle and per unit of apparent emitting surface). That this should be at all possible is only due to the fact that additional energy can be supplied to the beam of rays emitted and which illuminate the specimen, this being done by accelerating the electrons.

Thus it is seen that the high voltage does not serve the sole purpose of "hardening" the electrons to give them the power of penetrating through the specimen.

The high D.C. voltage is supplied by a cascade generator of a conventional type, consisting of a 50 kV transformer and a combination of two valves and two capacitors. The voltage is smoothed by a filter and adjusted to one of the four previously mentioned values by means of a variable ratio transformer preceding the high-tension transformer. A photograph of the generator which is mounted in an oil-filled casing is reproduced in *fig. 19*.

The position the generator occupies in the desk and its connection to the microscope tube by a rubber cable with earthed shield are to be seen in fig. 2. The cable runs behind the desk with a wide loop and is secured axially to the head of the microscope tube, so that sufficient play is left when the cathode end of the tube has to be opened for renewing the filament. The cable shield is connected air-tight to the earthed shield forming the head of the microscope tube — a construction similar to that found with "Metalix" X-ray tubes. Since there are no insulating surfaces exposed to the air, the instrument can be used even under unfavourable atmospheric conditions, and particularly under reduced atmospheric pressure (e.g. in laboratories high up above sea level), without any risk of flash-overs, via the insulation.

The focal length of a magnetic electron lens, energized by a current I, for electrons accelerated by a voltage V_0, is proportional to V_0/I^2 (at least so long as the iron of the pole pieces is not saturated; see formula (4) in the article quoted in footnote [1])). In order to get perfectly sharp micrographs it is therefore necessary that during say 30 seconds — about the longest time required for focusing an exposure — the accelerating voltage does not vary by more than 1/20,000 and the excitation currents by no more than 1/40,000.

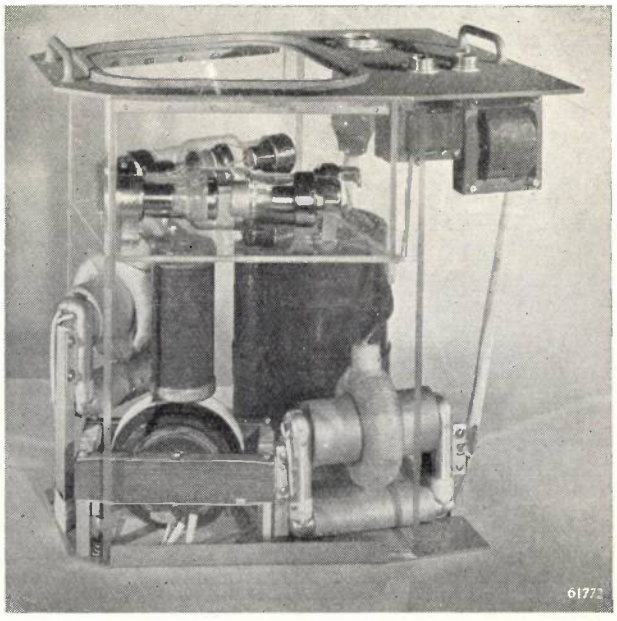


Fig. 19. The high-tension generator ready to be placed in the oil-filled tank. On the left at the bottom is the 50 kV high-tension transformer. On top of that is one of the capacitors and, still higher, the two valves for the cascade circuit (see the diagram in fig. 20); the second capacitor, together with the resistors and capacitors of the smoothing filter, is in the large cylinder on the right at the back. On the right at the front and on the extreme left, somewhat higher up, are the 100 kV insulated filament-current transformers for the cathode of one of the valves and for the filament of the electron gun.

There are various methods for stabilizing the high tension to the required degree. The most effective method consists in feeding back the fluctuations in the high tension to the high-tension generator [3]). In this way both the variations arising at the input (mains voltage fluctuations) and those arising at the output are counteracted. The output variations may be due, for instance, to reduced emission of the filament, variations in the filament voltage, release of traces of gas in the tube, or — to put it in general terms — to variations of the load. In our case, thanks to the connection of electron gun resembling a self-biased triode, the load of the generator is constant to a high degree, it thus being sufficient to stabilize the input voltage of the generator.

To this end, the generator is fed from a small valve oscillator (output 20 W) oscillating at a frequency of 100 c/s. The D.C. voltages required for feeding this oscillator are derived from a highly stabilized supply unit of known design [4]). *Fig. 20* is a circuit diagram of the complete high-tension generator.

The point gained in the method of stabilization employed is that it costs little trouble to keep a low D.C. voltage highly constant, and thus a highly constant A.C. output of the oscillator is ensured. Attempts to stabilize direct the alternating voltage of the mains and to use this as input for the high-tension generator, for instance by means of the known stabilizers with transductors, were doomed to failure in our case, because these stabilizers have too much inertia to smooth out surges and are also too sensitive for mains-frequency fluctuations nowadays so often occurring.

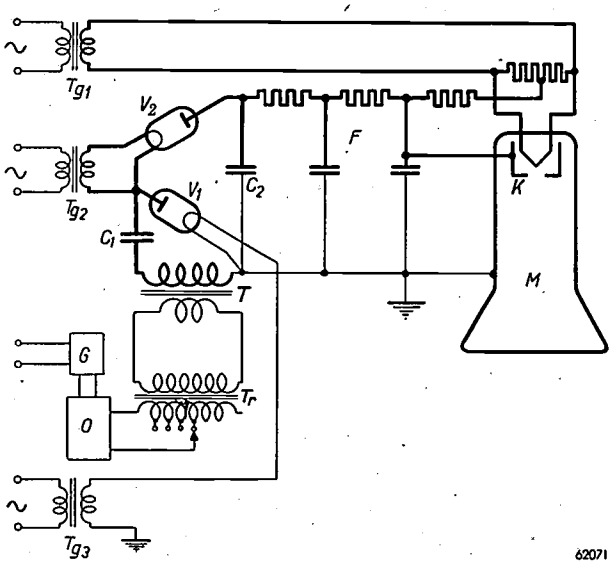


Fig. 20. Circuit of the high-tension generator. T high-tension transformer, C_1, C_2 capacitors and V_1, V_2 valves of the cascade circuit, T_r variable-ratio transformer, T_{g1}, T_{g2}, T_{g3} filament-current transformers, F smoothing filter, M microscope tube with cathode K and earthed anode. The generator is fed from the valve oscillator O (frequency 100 c/s) which receives the necessary anode and filament voltages from the stabilized supply unit G. The high-tension parts are drawn in heavy lines.

The choice of 100 c/s for the valve oscillator made it possible to use a high-tension transformer of a normal type frequently used in X-ray apparatus. With higher frequency it would have been necessary to have a transformer of a special low-capacitance construction.

Wherever accelerated electrons come into collision with matter they generate X-rays. This is especially the case on the condenser aperture, which receives roughly half the electrons emitted. The metal jacket of the microscope tube and the iron jackets of the lens coils make it unnecessary to take any special measures for protection of the user against the X-rays from this source. More attention had to be paid, however, to the X-rays generated on the fluorescent screen. These rays, it is true, only reach an appreciable intensity when there is the strongest possible concentration of the electron beam upon the specimen and the smallest magnifications are used, but this source of X-rays is very close to the observer's face. Consequently, in order to render this radiation harmless, the thick covering window of the microscope tube has been made of lead glass and the shield round the projection chamber lined with lead. The X-ray dose that may possibly be received during an 8-hour working day is therefore no more than 1/100th to 1/10th of the tolerance dose permitted in the various counties.

Energizing of the electron lenses

As already pointed out, the windings of the magnetic lenses have to be energized with a current that does not vary by more than 1 : 40,000 during 30 seconds. A stabilized supply unit like that mentioned above (see footnote [4])) cannot supply this energizing current direct; it supplies a constant voltage, whilst the resistance of the lens windings changes rather considerably with the temperature (in spite of the water cooling, the temperature is by no means to be regarded as constant). By modifying the circuit described in the article quoted [4]), however, a power pack can be constructed

[3]) See, e.g., A. C. van Dorsten, Stabilization of the accelerating voltage in an electron microscope, Philips Techn. Rev. **10**, 135-140, 1948.

[4]) H. J. Lindenhovius and H. Rinia, A direct current supply apparatus with stabilized voltage, Philips Techn. Rev. **6**, 54-61, 1941.

which yields a constant current. The system adopted by us is represented in a simplified form in *fig. 21* and explained in the caption. Roughly it can be said that the system has the tendency to keep the voltage constant between the points

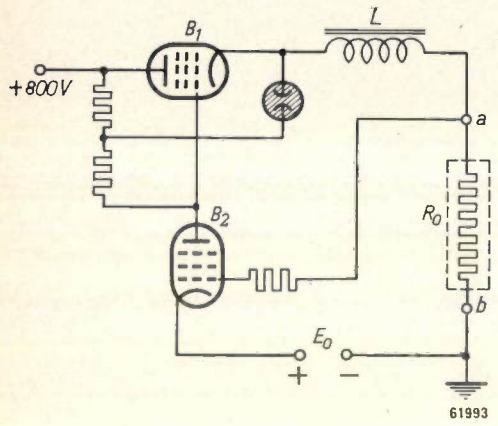


Fig. 21. Simplified circuit of the power pack supplying a constant current for an electron lens. The voltage drop across the constant resistor R_0 increases with the current flowing through the lens coil L, so that the control grid of the pentode B_2 (type EF6) becomes less negative, the current through B_2 increases, its anode potential drops and thus the control grid of the pentode B_1 (type EL51) becomes more negative, thus counteracting the tendency of the current through B_1 (= current through L) to rise. — The greater the amplification factor of the two valves and the higher the resistance of R_0, the more sensitive is the regulation and thus the smaller the resulting current fluctuations. In order to obtain the required grid bias for B_2 at a high value of R_0, the voltage drop across R_0 is compensated by a sufficiently high opposed voltage E_0 ($\approx$ 75 V). This voltage, which acts as a reference voltage, must of course be highly constant.

a and b and, since there is an accurately adjusted resistor R_0 with an extremely low temperature coefficient between those points, the current flowing through this resistor, which is the same as the current for energizing the lens winding L, is likewise kept highly constant.

The current can be adjusted to the desired value by varying the resistor R_0. This implies that a separate power pack according to fig. 21 is needed for each lens that has to be varied in strength independently of the others. This is the case with the diffraction lens and the intermediate lens respectively for varying the magnification, and also with the condenser lens for adjusting the "illumination". The focal lengths of the objective lens and the projector lens, on the other hand, the first of which has to be only slightly variable for focusing, can be coupled together. In all, therefore, we have three power packs. These, together with the common, small, stabilized unit supplying the constant reference voltage (E_0 in fig. 21), are shown in *fig. 22*.

The objective lens and the projector lens are interlinked by connecting the windings of these two lenses in series, with counterwise direction of the current. This has the advantage that the image rotations caused by the two lenses (see the caption of fig. 3) practically compensate each other. During focusing, therefore, no trouble is experienced from image rotation.

Also when varying the magnification, by adjusting the energizing current for the intermediate or diffraction lens, there is no image rotation worth mentioning, thanks to these lenses working with only a relatively small enlargement and thus in themselves causing hardly any image rotation.

When varying the accelerating voltage V_0 of the electrons it is necessary to change also the energizing currents for all lenses in order to get the same focal lengths (the variation required is inversely proportional to $\sqrt{V_0}$; see above). Each of the three power packs therefore has, for the four steps of the high tension, four different resistors R_0 which are automatically switched over when selecting the value of the high tension (with the large switch knob seen on the front of the desk in fig. 2). When the voltage is changed there is therefore hardly any need, if at all, to readjust the lenses.

The path followed by the electrons must not be

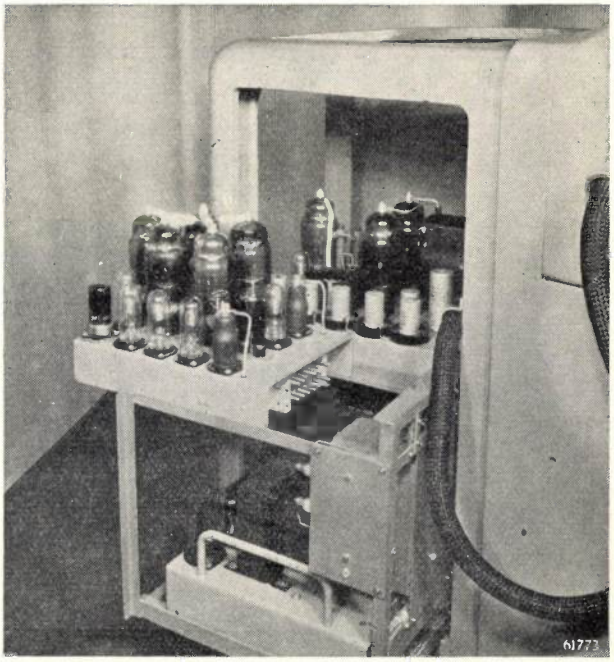


Fig. 22. Drawn-out chassis with the three power packs for the condenser lens (max. 100 mA), the diffraction or the intermediate lens (max. 100 mA) and the objective and projector lenses (max. 400 mA). Owing to the higher currents required for the last two lenses this power pack has three EF 51 valves connected in parallel, instead of the one pentode B_1 (fig. 21). At the back is the stabilized supply unit for the constant reference voltage E_0.

influenced by foreign magnetic (or electric) fields. In some microscopes this has had to be specially provided against, but in our microscope this has not been necessary because the lens coils with their iron jackets directly joining up with each other already form a very effective screening.

The vacuum system

To be able to work the microscope tube with 100 kV the pressure inside the tube must be lower than 10^{-4} mm Hg. Since, when working with the microscope, the vacuum is repeatedly disturbed and has to be restored, a rapidly acting and easily operated pumping system is essential. The system employed here is represented in *fig. 23*. It comprises

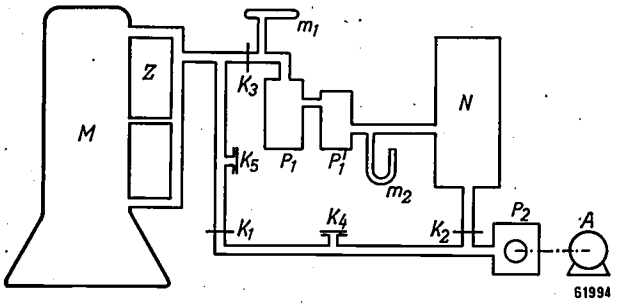


Fig. 23. Schematic diagram of the vacuum system. *M* microscope tube, *Z* pumping line, P_1, P_1' high-vacuum pumps, P_2 fore-vacuum (rotary oil) pump driven by the motor *A*; the high-vacuum pumps are water-cooled, the cooling system being connected in series with that for the electron lenses. *N* fore-vacuum tank, m_2 manometer for recording the fore-vacuum, m_1 Penning manometer for the high vacuum; the neon glow-lamp serving as indicator for the latter manometer can be seen through a slit in the front panel of the microscope desk (fig. 1). K_1-K_5 valves.

two high-vacuum pumps in series, viz. an oil-diffusion pump and a small mercury-vapour jet pump serving as a booster pump, capable of pumping at a rate of about 10 litres/sec at 10^{-4} mm Hg, and further a fore-vacuum pump (rotary oil pump) and a fore-vacuum tank. The pump line is connected to the microscope tube in three places: two wide connections are made to the cathode and projection chambers and a narrower one at the level of the selector diaphragm (see fig. 2). Thus when the air is being pumped out it does not have to travel through the whole of the microscope tube with its narrow passages. The connections are sealed off with rubber rings, in such a way that any vibrations from the high-vacuum pumps are not transmitted to the microscope tube.

The microscope is evacuated in four stages:

1) the microscope tube is pre-evacuated, the high-vacuum pumps being pre-heated;
2) the fore-vacuum tank, with the permanently connected high-vacuum pumps, is pre-evacuated;
3) the microscope tube is connected to the high-vacuum pumps;
4) the fore-vacuum tank is disconnected from the rotary oil pump and the latter stopped (at the same time air is admitted above the oil in this pump to prevent the oil from rising and entering the pump line).

Under normal conditions the whole cycle, starting with the microscope filled with air under atmospheric pressure, takes about 2 minutes [5]). When stage 4 has been reached, the fore-vacuum tank and the specimen lock enable the microscope to be used for several hours on end and a large number of specimens can be examined before it becomes necessary to switch on the rotary oil pump again for a few minutes.

The various necessary connections between the pumping lines are made with valves, which are easier to operate than the classical cocks. The lines have been so arranged that all these valves could be contained in one single valve box. All the connections that have to be made according to the above system are brought about by turning one single knob, driving a cam shaft inside the valve box which has a cam of appropriate profile for operating each valve as required.

In addition to the four positions corresponding to the four stages of pumping mentioned above, the valve-operating knob has a fifth position, the zero position (0) in which the microscope is shut off from all the pumps and atmospheric air can be let into the microscope by means of an inlet valve operated with a push-button. A filter in this valve line keeps out moisture and dust.

Operation of the instrument; safety measures

In the foregoing description of the Philips electron microscope stress has been laid on the measures taken for facilitating the operation of this instrument: the centering of the electron beam, the focusing of the image, checking of the apertures, insertion of the specimen, making micrographs, replacement of the filament, single-knob operation of the vacuum system.

The principle of single-knob operation and automatic control is applied in more than one connection in this microscope. Mention has already been made of the voltage selector, which automatically changes

[5]) This cycle always has to be passed through after a new film has been put in. Then, too, the microscope can be evacuated in two minutes, but only if the film has been stored in a dry place. If this simple precaution has not been taken then the evacuation may take hours, owing to the moisture absorbed by the gelatine layer of the film being only very slowly released.

the lens currents at the same time. Another example is the adjustment of the magnification: one knob covers the whole range of magnifications, from 1000 times to 60,000 times. The switching over from the diffraction lens to the intermediate lens, necessary for changing over from enlargements smaller than 4000 times to stronger magnifications, is done automatically with the same knob. Yet another example is the focusing device. Deflection is started simply by pressing a button, which is kept pressed in so long as one is adjusting the objective lens current. Finally, there is again the vacuum system: the spindle of the knob already referred to is also fitted with a control drum switching on the motor of the rotary oil pump in the positions 1, 2 and 3 and switching it off in the positions 4 and 0.

A number of measures have been taken to safeguard the instrument against damage through breakdowns or faulty manipulations. The control drum just referred to, coupled to the operating knob for the vacuum system, has an additional contact operating a relay in such a way that high tension can only be applied to the microscope tube when that knob is in position 4. Thus it is impossible for the high tension to be switched on by mistake before the tube has been evacuated, which would cause a gas discharge in the tube. If, however, the manometer should not have been properly watched and the knob prematurely turned to position 4, or if through some fault or other the vacuum in the tube should fail, nothing can even happen then, because an overload switch in the earth lead of the high-tension circuit trips the high tension as soon as the emission current in the tube exceeds 3 mA, a value corresponding to the occurrence of a glow discharge which does no damage to the microscope tube nor to the generator. Further, the knob for the vacuum system is fitted with a locking device which permits of its being fully turned in one direction only: for a complete cycle, from the air being admitted up to the tube being evacuated again, the knob has always to be turned from the zero position (0) through the positions 1, 2, 3 and 4 back to 0, in that order. If it were possible for the knob to be turned by mistake from 0, after air has been admitted into the tube, the wrong way round direct to position 4, in which the tube is connected to the high-vacuum pumps, then — owing to the large amount of air that they would suddenly be called upon to cope with — these pumps would be thrown out of proper working order for some hours. (For the same reason the push button used for admitting air into the tube is locked, so that it can only be pressed in in the zero position.)

Then there is also to be mentioned the safety switch fitted in the outlet of the water-cooling system. In the event of a stoppage in the circulation of the water the excitation of the electron lenses is automatically switched off, and also the heating of the high-vacuum pumps, so that no mercury vapour or oil vapour can get into the microscope tube. Simultaneously the heating of the high-vacuum pumps is automatically switched off — and then it is also impossible to switch on the high tension — if through some cause or other the fore-vacuum, measured by the mercury manometer m_2 in fig. 23, is insufficient. This safety device works with a relay which is switched on as soon as the mercury column of the manometer reaches a contact fitted in at a certain level.

Pilot lamps on the operating panel show at a glance what the situation is as regards the high tension, the water cooling and the fore-vacuum; instruments, some of which have already been mentioned, register the pressure inside the microscope tube, the magnification setting, the emission current, the filament current and the exposure time required, whilst there is also a check on the stabilization of the high tension.

This lengthy description may well have demanded some patience on the part of the reader, and also the impression that must have been gained from our introduction, namely that an electron microscope is not exactly a simple instrument, will no doubt now be more deeply rooted. It is well to realize, however, that the complexity of this electron microscope is for a large part not to be ascribed to the principle of the instrument and certainly does not involve proportional problems for the user. On the contrary, this seeming complexity results from the aims of the designers to relieve the user as far as possible of the burden of having to pay so much attention to this and that and from the necessity of having to make so many manipulations. As in so many other cases in modern engineering, part of the human work is transferred to the machine so that the operator can concentrate more on the possibilities that the machine offers him.

Summary. In the Philips electron microscope the microscope tube proper is mounted at an angle on a desk. The image is produced on a fluorescent screen 20 cm in diameter. Mounted in the desk is the high-tension generator, the pumping system and the supply apparatus for the electron lenses. The electron-optical system, based on principles developed by Le Poole, consists of five magnetic lenses. In addition to the usual condenser, objective and projector lenses there is an intermediate lens and a diffraction lens, by means of which the magnification can be continuously varied from 1000 times to 60,000

times, with the screen entirely covered, and one can change over immediately from the normal image of a specimen to an electron-diffraction diagram.

The microscope tube, with all the lenses, forms a rigid and accurately adjusted unit. The electron beam is centered in the tube with the aid of a set of deflection coils. The objective lens is so constructed that the object plane lies in the gap between the pole pieces (thus not in the bore). The specimen, as also the objective aperture, can therefore be moved into place from the side by means of a simple rod-like holder. A kind of air lock limits the amount of air entering the microscope to only a few mm^3 when exchanging specimens, and it takes no more than 20 seconds to restore the vacuum. The objective aperture and other diaphragms can easily be checked for centering and examined for contamination, and if necessary readjusted and cleaned. Neither is there any trouble in replacing the previously centered filament in the electron gun.

The images are photographed on standard 35 mm film. The sliding camera, with a film spool for 40 photographs, is within the vacuum. The camera is moved into the electron beam and the shutter opened from the outside, the film being automatically transported further. Focusing of the image is facilitated by a "wobbling" beam device.

The lenses are cooled with water and energized by three power packs supplying stabilized currents of maximum 100 and 400 mA. The accelerating voltage can be varied in four steps from 40 to 100 kV, the excitation currents of the lenses being thereby automatically adjusted so as to retain the image. The high-tension generator load is kept highly constant owing to the self-biased electron gun. Thus the necessary stabilization of the high tension is made possible merely by keeping the A.C. input voltage highly constant, this being achieved by employing a valve oscillator for the supply.

The fore-vacuum and high-vacuum pumps are operated in the right sequence by one manifold control knob. Various locking devices and safety switches safeguard the instrument against damage through breakdowns or faulty manipulations. Working with the microscope is further facilitated by various means, such as a device for uniform scanning of the specimen and for finding any part of the specimen again, an exposure meter, a marking device for indicating the scale on each photograph, etc.

AN EXPERIMENTAL "STROBOSCOPIC" OSCILLOSCOPE FOR FREQUENCIES UP TO ABOUT 50 Mc/s

I. FUNDAMENTALS

by J. M. L. JANSSEN.

621.317.755:621.3.029.5/.6

The highest frequency for which a sensitive cathode-ray oscilloscope of conventional design can be constructed is limited by various circumstances to about 10 Mc/s. An artifice makes it possible, however, to reach much higher frequencies. This artifice consists in the application of an electrical analogy of stroboscopic exposure, the commonly known method for studying periodical movements taking place so quickly that the eye cannot follow them. With the oscilloscope described here the stroboscopic flashes of light are replaced by electrical pulses which scan, as it were, the voltage curve that is to be examined and convert it into a phenomenon of low periodicity, which can then be viewed on the screen of a cathode-ray tube in the normal way.

Highest frequency attainable with oscilloscopes of conventional design

Cathode-ray oscilloscopes of the conventional type consist of a cathode-ray tube with electrostatic deflection, an amplifier for the voltage to be examined (vertical deflection), a device producing a sawtooth voltage for the time base (horizontal deflection) and some other parts with which we are not concerned here. Each of the three component parts mentioned sets a limit upon the highest frequency that can be properly observed in the oscillogram.

If, as the frequency is increased, the duration of a cycle of the deflection voltage becomes comparable to the time taken by the electrons to traverse the space between the deflection plates, a transit-time effect occurs in the cathode-ray tube. With an accelerating voltage of, say, 1000 V this effect becomes noticeable at frequencies of the order of 100 Mc/s.

The amplifier sets a much lower limit. If the amplification is to be kept independent of the frequency also at high frequencies, the amplification per stage has to be reduced to a low level, so that in order to get good sensitivity a large number of amplifying stages are required [1]). Moreover, there is a still more stringent requirement to be made of the amplifier for an oscilloscope, namely that the components of a non-sinusoidal voltage must be faithfully reproduced not only in amplitude but also in phase. Furthermore the amplifier must be capable of supplying a considerable reactive current. In fact, at a frequency of, say, 10 Mc/s the anode capacitance of the output valve and the parallel capacitance of the deflection plates together form an impedance of only something like 1000 ohms, so that with a voltage of, say, 100 V between the plates the capacitive current amounts to about 100 mA. A frequency of 10 Mc/s can therefore be regarded as being roughly the upper limit for which an oscilloscope amplifier can be built that answers reasonable requirements.

Together with the frequency of the voltage to be examined there is, of course, also the frequency of the time base to be raised. The highest frequency at which a sufficiently large sawtooth voltage can be generated with reasonable linearity and with short flyback time is about 1 Mc/s. If the frequency of the voltage to be examined (the voltage across the pair of vertical deflection plates) is, say, 10 Mc/s then the oscillogram never comprises less than about ten cycles, a number which for various purposes is more than is desirable.

In the normal Philips oscilloscopes so far produced lower frequency limits than those mentioned have had to be chosen in order to avoid having to make these oscilloscopes for a wide range of applications unnecessarily heavy and expensive. Thus the maximum frequency for the vertical deflection of the type GM 3152 oscilloscope [2]) is 1 Mc/s and that of the type GM 3159 [3]) 0.5 to 1 Mc/s (according to the sensitivity), whilst for both these types the highest time-base frequency is 150 kc/s. There are other types where, for special reasons, these limits are still lower.

Naturally there is also a need for an oscilloscope

[1]) See e.g. H. J. Lindenhovius, G. Arbelet and J. C. van der Breggen, A millivoltmeter for the frequency range from 1 000 to 30 × 10^6 c/s, Philips Techn. Rev. 11, pp. 206-214, 1949/1950 (No. 7).

[2]) Philips Techn. Rev. 4, 198-204, 1939.

[3]) Philips Techn. Rev. 9, 202-210, 1947.

for frequencies higher than 1 Mc/s. Instead of developing such an oscilloscope along the traditional lines, offering but little prospect of reaching frequencies higher than 10 Mc/s, we have found it preferable to strike out in an entirely different and more promising direction. In view of the resemblance it shows to the principle of stroboscopic exposure the oscilloscope that has now been devised has been called the "stroboscopic" oscilloscope.

This first article explains the fundamentals of the stroboscopic method, whilst a further article will be devoted to the electrical design of the various components.

Principle of the stroboscopic oscilloscope

When positive voltage pulses of constant amplitude are fed to the anode of, for instance, a pentode (*fig. 1*), whilst there is only a (negative) D.C. voltage on the control grid (this voltage not cutting the valve off), then the anode current consists of pulses of an amplitude depending upon the control-grid voltage. When in addition a small alternating voltage v_o is applied to the control grid and the frequency f_o of that voltage is equal to the repetition frequency f_i of the pulses on the anode (*fig. 2a*), then the anode-current pulses all have the same amplitude, this being determined inter alia by the phase of the pulse with respect to the A.C. grid voltage. The same thing holds when f_o is an exact multiple of f_i (fig. 2*b*). But if f_o deviates somewhat from f_i, or from nf_i (where n is a whole number), then the phase of each pulse with respect to the alternating voltage differs from that of the preceding one (fig. 2*c*). The pulses then lag or lead with respect to the alternating voltage and thereby scan that voltage point by point. Thus the anode-current pulses vary in amplitude and form, as it were, a series of snapshots of the amplitude of the alternating voltage. The lowest frequency f_z occurring in the anode current equals the absolute value of the difference between f_o and nf_i. If the voltage to be examined contains m harmonics of frequencies $2f_o$, $3f_o$, mf_o then the frequencies $2f_z$, $3f_z$, mf_z also occur in the anode current.

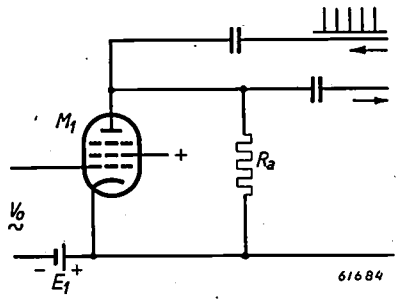


Fig. 1. Circuit of a mixing valve (pentode M_1) in which the alternating voltage v_o to be examined is mixed with voltage pulses fed to the anode. R_a anode resistor from which the output voltage is taken.

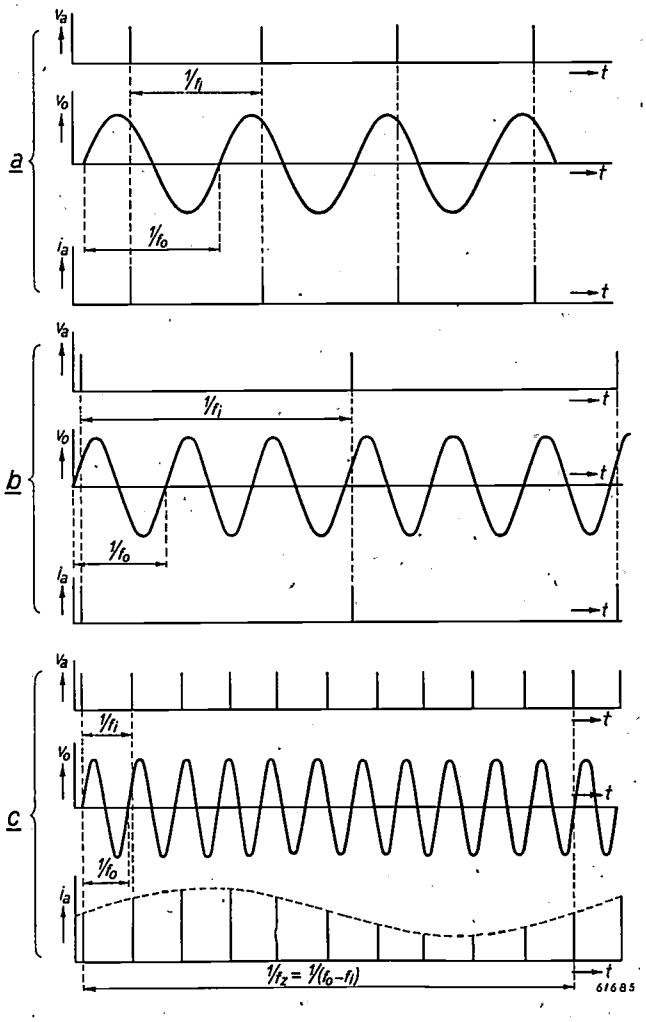


Fig. 2. Diagrammatic representation of the mixing process, representing as functions of the time t: v_a = voltage pulses applied to the anode, v_o = alternating voltage on the control grid of the mixing valve, i_a = anode current pulses. At (*a*) the frequency f_o of v_o is equal to the repetition frequency f_i of the pulses, at (*b*) $f_o = 3\ f_i$ and at (*c*) there is a small difference between f_o and f_i. In the last case the series of anode-current pulses forms a point-by-point reproduction of v_o and the frequency $f_z = f_o - f_i$ (generally $|f_o - nf_i|$) takes the place of f_o.

The analogy with stroboscopic exposure becomes evident here: in the two cases first mentioned ($f_o = f_i$ and $f_o = nf_i$) the pulses correspond to synchronous flashes of light, whereby the exposed, periodically moving object (frequency f_o) is made to appear to be quite stationary; in the case corresponding to that where f_o differs somewhat from nf_i on the other hand the object appears to be moving slowly. It is this latter case with which we are concerned in the designing of the "stroboscopic" oscilloscope: the amplitudes of the anode-

current pulses form "reproductions" (albeit only of a finite number of points) of the alternating voltage at the control grid, and the fundamental frequency of the anode current, $f_z = |f_o - nf_i|$, can be made much lower than f_o. It is in this latter possibility that the essential advantage of the stroboscopic method lies for those cases where f_o is beyond the range of an ordinary oscilloscope.

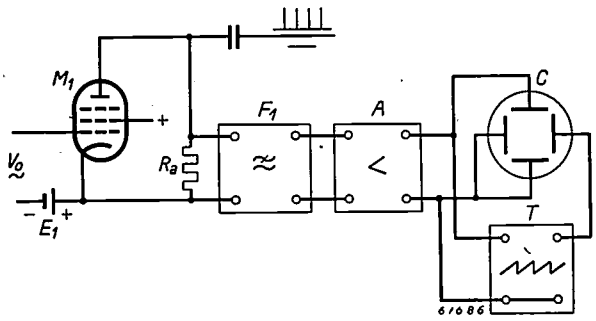


Fig. 3. Block diagram of a stroboscopic oscilloscope with uniform scanning. v_o the voltage to be examined, M_1 mixing valve with an anode resistor R_a, F_1 low-pass filter, A low-frequency amplifier, C cathode-ray tube, T generator of a sawtooth voltage the frequency of which is synchronized with the frequency f_z taken from the amplifier A.

A stroboscopic oscilloscope could in principle be devised in the following way: a resistor R_a (fig. 1 and *fig. 3*) is inserted in the anode circuit of the pentode (from now on we shall call this the mixing valve, since in this valve the voltage v_o is mixed with the impulses [4])); with the aid of a low-pass filter only the low frequency components (f_z and its multiples) are extracted from the voltage across R_a, this filter output then being amplified and fed to the vertical deflection plates of a cathode-ray tube.

To give the oscillogram a linear time scale the voltage for the horizontal deflection has to be given the shape of a linear sawtooth voltage, since the phase of the scanning pulse with respect to the alternating voltage to be examined increases linearly with time. Furthermore, in order to get a stationary picture the sawtooth voltage has to be given the frequency f_z (or a frequency of which f_z is a multiple).

The repetition frequency f_i of the pulses has to be so adjustable that a multiple of this frequency, nf_i, can be brought close enough to f_o to give $f_z = |f_o - nf_i|$ a value lying below the cut-off frequency of the filter. And this has to be the case for all values of f_o between the widest possible limits; thanks to the factor n, which can be chosen at will, f_i only needs to be adjustable within a limited range. If v_o contains harmonics up to and including the m^{th}, then mf_z must also be passed by the filter.

Although it is in principle possible for an oscilloscope to be arranged along these lines, great difficulties would be encountered in its execution. With f_o, say, = 30 Mc/s, and thus nf_i differing but little from that value, the variation in time of f_i would have to be much less than can be realized in practice; even 0.1% variation of nf_i means a change of 30 000 c/s in f_z. As a consequence f_z, and still more so any multiples of f_z, would then come to lie above the cut-off frequency of the filter. Moreover, it would be impossible to keep the sawtooth generator of the time-base voltage synchronized with such a variable f_z.

With the method described below this difficulty is avoided and, furthermore, another possibility is opened.

Phase modulation of the pulses

With the method applied by us the pulses are periodically modulated in phase (or in position, if this term is preferred). This means a modulation of the repetition frequency f_i, the average (or central) value of which, f_{ic}, is so chosen that $nf_{ic} = f_o$.

On the average, therefore, the pulse generator is synchronized with the frequency f_o. Without phase modulation the pulses would continuously be scanning the same point of the voltage curve (fig. 2*b*), but by "moving them to and fro" — that is to say, by periodically modulating them in phase — they are made to scan different points of the v_o curve.

In place of the phase difference increasing linearly with time, as with the previous method, we therefore have here a phase difference varying according to a periodical function of time. It is not of primary importance what periodical function is chosen for this, but in order to have a linear time scale the voltage chosen for the horizontal deflection must be the same periodical function of time as that of the phase difference. For both functions one could use, for instance, one and the same sawtooth function, but it is simpler to use a sinusoidal function, especially if it is given the frequency of the mains; it will be shown presently why such a low frequency is advantageous. This is illustrated in *fig. 4*, where in the diagram (*a*) a cycle of the voltage v_o to be examined is represented (superposed upon a grid bias E_1) and in the diagram (*b*) the

[4]) The ratio of the amplitude of the anode-current component with frequency f_z to the amplitude of the fundamental wave of v_o will also be called here the conversion conductance.

scanning pulse is shown in the state of rest (without phase modulation). When the phase φ of the pulse is made to change sinusoidally with the time t, then upon swinging from $-\pi$ to $+\pi$ (fig. 4c)

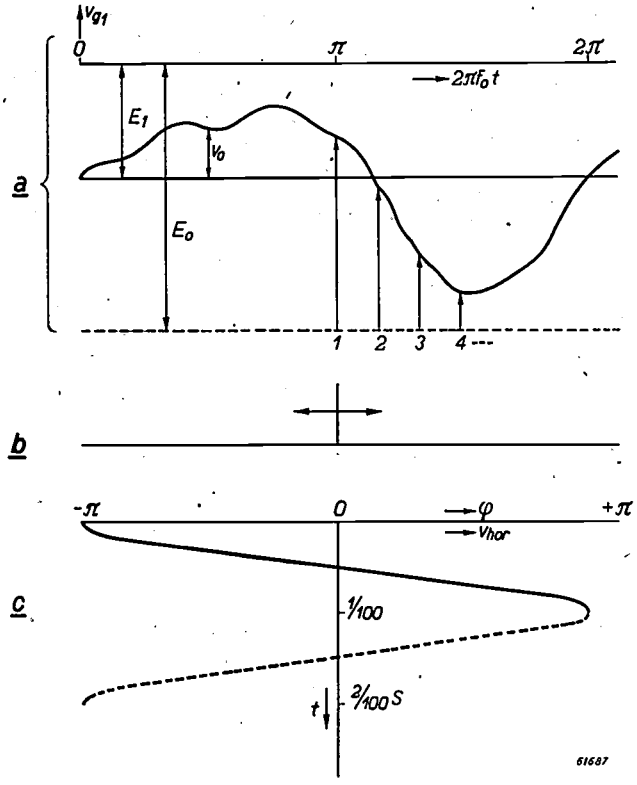


Fig. 4. Scanning with phase-modulated pulses. *a*) Grid voltage v_{g1} of the mixing valve, consisting of the bias E_1 and the voltage v_o to be examined. At the grid voltage E_0 the valve is just cut off. *b*) State of rest of the voltage pulse on the anode. *c*) Phase φ of the pulses swinging about the state of rest, and the voltage v_{hor} for the horizontal deflection, as function of the time t plotted vertically downwards. The frequency of φ and v_{hor} is the mains frequency (50 c/s).

the pulse scans a number of points of the v_o curve in succession. In fig. 4*a* the vertical arrows *1, 2, 3*... — drawn from the level E_0, i.e. the grid voltage at which the mixing valve is just cut off on the occurrence of a pulse — indicate the amplitude of several successive anode-current pulses, which are the "snapshots" of the v_o curve. The low-frequency components of the anode current, with the frequency of the phase modulation and a series of multiples of that frequency (inherent in the frequency spectrum of the phase modulation), bring about the vertical deflection of the electron beam in the cathode-ray tube. The horizontal deflection is brought about by v_{hor} (fig. 4*c*) varying synchronously with the movement of the pulse. The screen of the cathode-ray tube then shows a curve which — provided the instantaneous reproductions are sufficient in number — is a faithful picture of one cycle of the v_o curve.

During the return movement of the pulse (φ varying from $+\pi$ to $-\pi$ according to the dotted part of the sinusoidal line in fig. 4*c*) the pulse could be made to scan the same v_o curve again, the spot of light on the screen then describing the same curve as before but in the reverse direction, since v_{hor} then changes in the opposite sense. Better use can be made of the return stroke by causing the pulse to scan another voltage curve, so that the oscillograms of two voltages (v_{oI} and v_{oII}) can be obtained simultaneously. (v_{oII} may possibly be zero, in which case the spot of light describes the zero line on the return stroke.) We shall see in a subsequent article how the alternate scanning of two curves can be brought about with the aid of a simple electronic switch.

In fig. 4 the amplitude of φ (the "phase sweep" of the pulse) has been made equal to half a cycle of v_o, so that exactly one whole cycle of the v_o curve is scanned. There is nothing, however, to prevent the sweep being made larger or smaller so as to be able to scan a part of the curve that is more or less than one cycle. Since this scanning takes place in the same interval of time (1/100th sec) as that in which the horizontal movement completes one half cycle, the part scanned always covers the full width of the oscillogram. By giving the pulse a small sweep and making the state of rest around which it sweeps adjustable, any part of the v_o curve can be observed as it were microscopically. This is of great value for studying a certain detail, a possibility which does not exist when applying the method previously mentioned (pulses proceeding at constant speed along the v_o curve).

Fig. 5 shows how a stroboscopic oscilloscope with sinusoidal phase-modulated pulses can be designed. (Although it is not customary to use phase modulation of the light flashes in actual

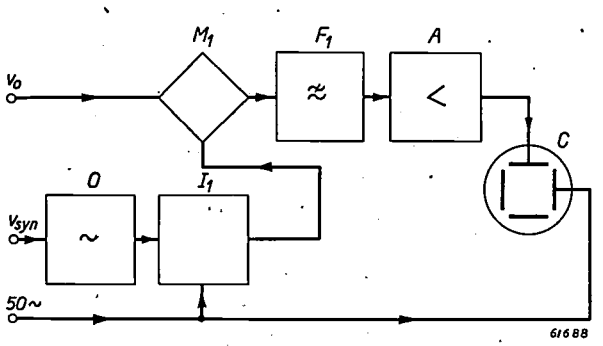


Fig. 5. Elementary block diagram of a stroboscopic oscilloscope with sinusoidal phase modulation. v_o, M_1, F_1, A and C as in fig. 3. I_1 pulse generator a harmonic of which is synchronized via the oscillator O with the synchronization voltage v_{syn} (synchronous with v_o). I_1 is phase-modulated by the mains voltage.

stroboscopy the term "stroboscopic" is nevertheless used here.) As was the case with the system shown in fig. 3, vertical deflection is brought about by the output voltage of the mixing valve M_1 via a filter and an amplifier. An oscillator O, supplying a sinusoidal voltage, is synchronized with an externally applied voltage v_{syn}, which has to be synchronous with v_o (v_{syn} may in fact be the voltage v_o itself). The output voltage from this oscillator synchronizes in turn a pulse generator I_1, to which in addition a sinusoidal voltage v_φ of 50 c/s is applied for modulating the pulse in phase. The frequency of the oscillator is the same as the repetition frequency of the pulses, for which we have chosen — for reasons which will be explained later — a central value f_{ic} of approximately 100,000 c/s; f_{ic} is so adjusted that a multiple of it is just equal to the frequency f_o.

These components will be described in further detail in a subsequent article.

Limitations of the stroboscopic oscilloscope

We shall now first investigate what limitations have to be considered in the case of a stroboscopic oscilloscope and how far these set limits to the uses of such an apparatus.

In the foregoing it has been shown that the "data" of the oscillogram consist only of a finite number of points (the peak values of the anode-current pulses; see, e.g., fig. 2c). The question is in how far these points are able to give a faithful picture of the original v_o curve in spite of the gaps in between them.

The second point that we have to consider more closely is the duration of the pulses. So far we have tacitly assumed it to be infinitely small. Of course such pulses cannot be realized, and we must therefore work with pulses of a finite duration. Just as the finiteness of the duration of a stroboscopic flash of light causes a certain kinetic blurring of the object observed, so with a finite duration of the electrical pulse certain details of the v_o curve become lost; in other words, the resolving power of the pulses with which we have to work is not unlimited.

We shall now deal with the question of the gaps between the pulses and then with that of the duration of the pulses.

Gaps in the oscillogram

If the v_o curve contains m harmonics then it is defined by $2m + 1$ points of a cycle, since this number of points determines the $2m + 1$ coefficients of the Fourier series, namely m coefficients of the sine terms, m coefficients of the cosine terms and the constant term (D.C. voltage component). Given that the cycle of the curve has p points, then — notwithstanding the gaps between these points — the curve is completely determined if $p \geqq 2m + 1$, or

$$p > 2m \quad \ldots\ldots\ldots \quad (1)$$

If this condition is not satisfied the best approximation that can be derived from the inadequate data is the curve having no more than $m' = \frac{1}{2}p < m$ harmonics corresponding to the data. In order to reproduce with a stroboscopic oscilloscope as many harmonics as possible this apparatus has to be so designed that the picture consists of the largest possible number of points. This means that the rate of scanning (the speed at which the pulses move along the v_o curve) has to be as low as possible.

To work this out quantitatively let us first assume that we again have to do with the case with which we began our considerations about the stroboscopic oscilloscope, namely that of a constant repetition frequency f_i of the pulses of which the n^{th} harmonic differs somewhat from the fundamental frequency f_o of the voltage v_o. The rate of scanning is then constant, corresponding to the constant differential frequency $f_z = |f_o - nf_i|$ and with equidistant "measuring data" (anode-current pulses). It is easily verified that f_i/f_z is the number of points p with which one cycle of the v_o curve is scanned. Therefore, in order to observe even the m^{th} harmonic of this curve, provision has to be made to satisfy the condition

$$\frac{f_i}{f_z} > 2m. \quad \ldots\ldots \quad (2)$$

A similar condition holds for the method actually applied where the pulse is modulated in phase and consequently also in frequency. The "measuring data" are now no longer equidistant as in the case considered above, but for small instantaneous values of the phase sweep ($\varphi = 0$), where the pulse has a high scanning speed, these "data" lie farther apart than is the case with large instantaneous values ($\varphi = +\Delta\varphi$ or $-\Delta\varphi$), where the scanning speed is only low. In other words, the middle part of the oscillogram produced is built up from fewer data than the parts to the left or right of it. In order to get the same "density of measuring data" in this middle part as is obtained with uniform scanning, it is therefore necessary that the condition (2) shall be satisfied when the maximum value is substituted for f_z.

Since the frequency $\nu = 50$ c/s with which the phase modulation takes place is very low compared with $f_{ic} \approx 100{,}000$ c/s, we may regard the situation as being quasi-stationary and speak of the momentary repetition frequency f_i' of the pulses sinusoidally swinging about the central repetition frequency $f_{ic} = f_o/n$. We then have a variable differential frequency $f_z' = |f_o - nf_i'|$, for which, as will be derived below, we find the expression $|\nu \cdot \Delta\varphi \cdot \sin 2\pi\nu t|$ (where $\Delta\varphi$ is the phase sweep of the pulses expressed in radians of the v_o curve).

If $2\pi f_{ic}t + \psi$ is the phase of the unmodulated pulse (ψ being an arbitrary constant) then the phase of the modulated pulse can be written as $2\pi f_{ic}t + \psi + (\Delta\varphi/n)\cos 2\pi\nu t$, where $\Delta\varphi/n$ is the phase sweep expressed in radians of the series of pulses.

The momentary angular frequency of the pulses is defined as the derivative of the phase with respect to time [5]), i.e. $2\pi f_{ic} - 2\pi\nu \cdot (\Delta\varphi/n) \cdot \sin 2\pi\nu t$, and the momentary frequency f_i' itself is therefore

$$f_i' = f_{ic} - \nu \cdot (\Delta\varphi/n) \cdot \sin 2\pi\nu t \quad . \quad . \quad . \quad . \quad . \quad (3)$$

For the momentary differential frequency f_z' we then find $f_z' = |f_o - nf_i'| = |\nu \cdot \Delta\varphi \cdot \sin 2\pi\nu t|$, since $nf_{ic} = f_o$.

Substituting for f_z in (2) the maximum value of $f_z' = |\nu \cdot \Delta\varphi \cdot \sin 2\pi\nu t|$, i.e. $f_{z'\max} = \nu \cdot \Delta\varphi$, we get

$$\frac{f_{ic}}{\nu \cdot \Delta\varphi} > 2m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For the case where exactly one cycle of the v_o curve is scanned we must have $\Delta\varphi = \pi$ radians (cf. fig. 4c), so that (4) then becomes

$$\frac{f_{ic}}{2\pi\nu} > m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

From these formulae it appears that it is favourable to choose a high central repetition frequency f_{ic} of the pulses and a low frequency for the phase modulation (ν). As regards the latter it is an obvious solution to choose for ν the mains frequency; any lower frequency would have to be specially generated, whilst moreover there would soon be a flickering of the picture to contend with, since the time base also has to have the frequency ν.

With $f_{ic} \approx 100{,}000$ c/s (see the following section) and $\nu = 50$ c/s we find from (4a), for the scanning of a complete cycle, that the highest harmonic that can be observed is given by $m = 10^3/\pi \approx 300$. If less than one cycle is scanned ("microscopic scanning", $\Delta\varphi < \pi$) then the resolving power reaches to harmonics $\pi/\Delta\varphi$ times as high.

It might be thought that the formulæ (4) and (4a) hold only for the very special case where f_{ic} and thus also f_o are exact multiples of ν, since then with each sweep the same points of the v_o curve are scanned every time, and that in all other cases therefore, where different points are scanned every time, the said conditions need not be complied with. Such, however, is not the case. A closer analysis shows that it is a question of the number of points scanned by one sweep of the pulse, regardless of the question whether following sweeps cover the same points or different ones; it is thus independent of chance values of the ratio f_{ic}/ν. As a matter of fact the relations (4) and (4a) will presently be derived once more in a manner which shows this independency.

Duration of the pulse

In the foregoing section we have seen that with infinitely small pulses the v_o curve to be examined can only be displayed with harmonics of up to a limited order (m), independently of the fundamental frequency f_o. This limitation is due to the gaps. The following will show that actually the finite width of the pulse sets an absolute limitation upon the frequency for which a stroboscopic oscilloscope can be used. If the v_o curve contains harmonics of a frequency higher than this limit $f_{\max}$ then those harmonics are not reproduced (or at most but poorly), even though their order may not exceed the above-mentioned limit m.

When the oscillation time of the voltage to be examined becomes of the same order as the duration τ of the pulses, then the conversion conductance of the mixing process and thus the sensitivity of the instrument rapidly diminishes. The frequency limit $f_{\max}$ may therefore be said to be of the order of $1/\tau$.

To define this more precisely it has to be borne in mind that when the frequency is equal to $1/\tau$ the sensitivity is zero, at least when the pulses are rectangular (*fig. 6*), since at that frequency the

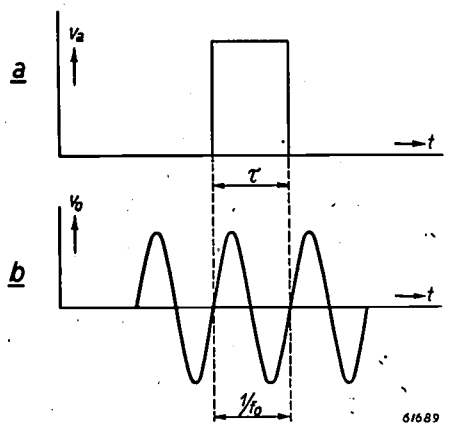


Fig. 6. Rectangular anode voltage pulse v_a of which the duration τ is exactly equal to one cycle $1/f_o$ of the voltage v_o to be examined.

[5]) See e.g. Th. J. Weyers, Frequency modulation, Philips Techn. Rev. 8, 42-50, 1946, and in particular page 44.

width of the pulse just matches one cycle of the voltage v_o, so that the contribution that the average value of the anode current of the mixing valve receives from the positive half of this cycle is exactly cancelled by the contribution from the negative half, and thus the conversion conductance is nil.

In a subsequent article we shall show that the pulses used have roughly the shape of a half sine. With this shape of pulse the rule also holds that the conversion conductance is nil at a frequency of $1/\tau$, if τ is understood as being a kind of average width of the pulse corresponding to about $^2/_3$ of the width at the base.

To allow for a sufficiently wide margin away from the state where the conversion conductance is zero one can take as the frequency limit

$$f_{max} \approx \frac{1}{2\tau} \quad \ldots\ldots\ldots \quad (5)$$

With a pulse duration of the order of 10^{-8} sec — in a subsequent article it will be shown what difficulties stand in the way of generating pulses of shorter duration — it follows that $f_{max} \approx 50$ Mc/s.

Using the results obtained in the preceding section we can arrive at an equation connecting the central repetition frequency f_{ic} of the pulses, the frequency ν of the phase modulation and the frequency limit f_{max} (or the pulse duration τ).

Suppose that the fundamental frequency of the voltage v_o to be examined is of the lowest value that can be considered for stroboscopic scanning, thus $f_o = f_{ic}$. Let us also assume that the highest (m^{th}) harmonic of v_o has exactly the frequency f_{max}. Then $m = f_{max}/f_{ic}$. And at the same time it is desired that the m^{th} harmonic is just the highest that can be reproduced in view of the gaps, so that the condition (4) holds. Let the phase sweep be such that, measured on the scale of the frequency f_{ic}, which always has about the same value, it amounts to π radians. In the case considered here ($f_o = f_{ic}$) the phase sweep is then likewise π radians on the scale belonging to f_o and thus just sufficient for scanning one cycle of the v_o curve; in other cases ($f_o > f_{ic}$, $n > 1$) more than one cycle can be scanned. The condition (4) now has to be applied in the form (4a), and with $m = f_{max}/f_{ic}$ this becomes:

$$\frac{f_{ic}}{2\pi\nu} > \frac{f_{max}}{f_{ic}}$$

or

$$f_{ic}^2 > 2\pi\nu f_{max} \approx \frac{\pi\nu}{\tau}.$$

From this equation it follows that, with τ in the order of 10^{-8} sec and $\nu = 50$ c/s, f_{ic} > about 100,000 c/s. Since the repetition frequency of the pulses forms the lowermost limit of the frequency range of the oscilloscope, in order to keep this range as wide as possible f_{ic} has not been chosen any higher than is necessary, hence about 100,000 c/s, the value frequently mentioned in the foregoing.

Cut-off frequency of the filter

The required cut-off frequency of the low-pass filter (F_1 in fig. 5) is directly related to the value of f_{ic}, as will be understood from what follows.

At (*a*) in *fig. 7* part of the frequency spectrum of the phase-modulated pulses has been drawn: it shows the fundamental frequency f_{ic} and some harmonics, $(n-1)f_{ic}$, nf_{ic} and $(n+1)f_{ic}$, which would already be present in the case of unmodulated pulses, and also the side bands due to the phase modulation; the distance between two adjacent lines of the side bands is the modulation frequency ν. At (*b*) in this diagram we have represented the single-line spectrum of the voltage v_o (any harmonics of v_o are of no consequence here); this line lies at $f_o = nf_{ic}$. Mixing of the two spectra produces innumerable differential frequencies (the differences between f_o and all frequencies of the spectrum of

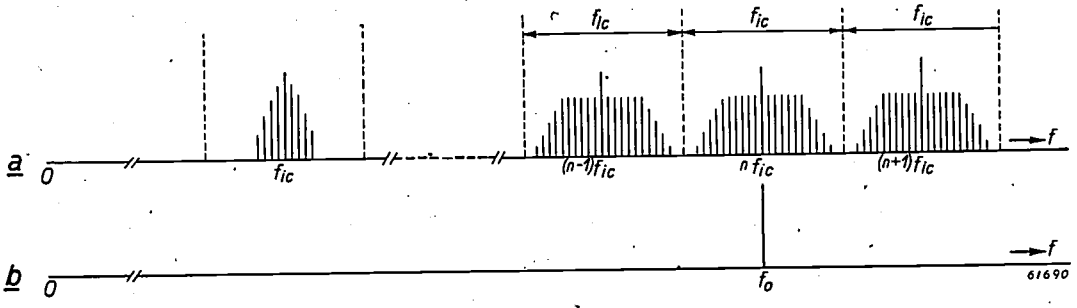


Fig. 7. Frequency spectrum, (*a*) of the phase-modulated pulses, (*b*) of the sinusoidal voltage v_o. (At (*a*) neither the number of components of the side bands nor the intensity of the components are true to scale; it is only a schematic indication that the side bands of the multiples of f_{ic} are richer in components than the side bands of f_{ic} itself, and that the outermost components of the side bands are much weaker than the majority of the components lying farther inwards.)

the modulated pulses). Of all these frequencies the filter should pass only those derived from f_o and the side bands of nf_{ic}; frequencies which are the difference of f_o and other side bands — in particular the adjacent side bands of $(n - 1)\, f_{ic}$ and $(n + 1)\, f_{ic}$ — have to be suppressed. From this it follows that the cut-off frequency of the filter must not exceed $f_{ic}/2 \approx 50{,}000$ c/s.

From these considerations it also follows that the side bands should not overlap, hence that their width should not be more than $f_{ic}/2$; in other words in fig. 7 they should not extend beyond the vertical dotted lines. Strictly speaking they actually do, because they consist of an infinite number of terms, but if, as is the case here, the situation is to be regarded as quasi-stationary ($\nu \ll f_{ic}$) the part of the side band falling outside the frequency sweep is of so small an amplitude as to be negligible.

The condition that has to be satisfied to ensure that the side bands do not overlap is, therefore, that none of the multiples of the pulse frequency may have a greater frequency sweep than $f_{ic}/2$. For the n^{th} harmonic (belonging to the fundamental frequency f_o) the frequency sweep is $\nu \cdot \Delta\varphi$, as follows from the equation (3) derived in the foregoing. If the v_o curve contains harmonics with m as the highest order then the corresponding multiple $mn \times$ (pulse frequency) makes the sweep $m\nu \cdot \Delta\varphi$. This is the largest of all frequency sweeps that have to be considered, and the fact that this must be smaller than $f_{ic}/2$ immediately leads to

$$\frac{f_{ic}}{2\nu \cdot \Delta\varphi} > m,$$

and, with $\Delta\varphi = \pi$ (scanning of one cycle of f_o), to

$$\frac{f_{ic}}{2\pi\nu} > m\,,$$

so that we have again arrived at the formulae (4) and (4a) by a different method.

The electrical design of a number of components of an experimental stroboscopic oscilloscope, such as the pulse generator, the system for synchronization, etc., will be dealt with in a further article.

Summary. A new principle is described for a cathode-ray oscilloscope which is very promising for the investigation of voltages with very high frequenties and shows a close connection with the known stroboscopic principle. According to this principle the voltage to be examined is mixed in a mixing valve with a pulsatory, phase-modulated voltage; the central repetition frequency of the pulses is so chosen that some multiples of it is exactly equal to the frequency of the voltage to be examined. The anode current of the mixing valve consists of pulses which produce, as it were snapshots in a low-frequency rhythm, of the voltage examined. The low-frequency components of the anode current are filtered out and used for the vertical deflection in the cathode-ray tube. Without phase modulation of the pulses these would "scan" the same point of the voltage curve every time, but by periodically modulating the pulses in phase they can be made to scan at will either a larger or a smaller part of the curve, depending upon the extent of the phase sweep. A stationary picture with linear time scale is obtained on the screen of the cathode-ray tube by arranging for the horizontal deflection to take place according to the same time function as that for the phase modulation. It is advantageous to choose for this time function a sinusoidal function having the mains frequency. Consideration is given to the influences of the gaps between the "snapshots" from which the oscillogram has to be reconstructed, and of the finite duration of the pulses used.

THE DISTRIBUTION OF ILLUMINATION ON A PLANE PARALLEL TO A TUBULAR LAMP

by J. LOEB *).

535.245.1:628.932.1

The lamps used prior to the invention of the tubular fluorescent lamp can almost invariably be regarded as point sources of light. Text books on illumination engineering show how the illumination fram such lamps can be determined in a given plane. Now that tubular lamps are becoming of more general use it is of importance to have a method by means of which the illumination in a given plane can easily be calculated for this kind of lamp too. With the method described here it is only necessary to consult a simply arranged table that is easily compiled.

Regarding the calculation of the illumination when using tubular lamps, two articles [1) 2)] have already been devoted to this subject in this journal. Several investigations dealing with the same problem have also been published in other periodicals; in particular mention is made of a treatise by K. Norden [3)] published in 1908.

Now that tubular lamps, and especially tubular fluorescent lamps, are becoming of more and more general use it is of importance that everyone having to make calculations of the illumination when using these lamps should have at his disposal the means for determining these illuminations in a simple manner.

Since the diameter of a tubular lamp is always small compared with its length, there is no objection to the lamp being considered, diagrammatically, as a line of light. In this article, therefore, we shall always speak of linear sources of light. In some investigations an approximative solution is considered sufficient, the linear source of light being regarded as a point source. In practice such an approximation suffices if the point for which the calculation has to be made is a reasonable distance away from the lamp (cf. [1)]), but for accurate calculation of the illumination at shorter distances from the lamp we shall avoid any such approximation. We shall assume, however, that the radiation from the lamp follows Lambert's law.

In this article we shall confine our considerations to a study of the direct illumination on a horizontal plane obtained from a linear source of light (or number of sources of light parallel to the plane) situated on a higher horizontal plane, for instance against the ceiling of the locality. This is the situation most commonly occurring in practice, where it is often desired to know what illumination is obtained on a horizontal table, for example in a workshop, from a general lighting at ceiling level or local lighting from lamps parallel to the plane of the table. This calculation can be made for lamps with and without fixture. It is also possible to determine in like manner the distribution of the illumination in any plane with respect to the lamp, allowance being made, moreover, to a sufficient approximation, for the light reflected from the ceiling and the walls of the location in which the lamp is installed [4)].

Lamp without fixture

Let us first consider the case where a linear source of light is used without fixture. Assume the lamp to be of the length l and mounted horizontally along the ceiling at a height h above the horizontal work plane. With the aid of *fig. 1* we shall now

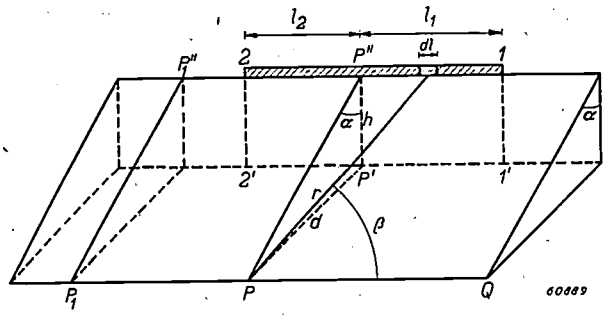


Fig. 1. Calculation of the illumination in a point P of a horizontal plane yielded by the element dl of a lamp fixed to the ceiling. *1* and *2* are the extremities of the lamp. The other figures and letters in the diagram are explaned in the text.

*) Lampes Philips S.A., Geneva.

1) N. A. Halbertsma and G. P. Ittmann, Illumination by means of linear sources of light, Philips Techn. Rev. **4**, 181-188, 1939.

2) H. Zijl, The calculation of lighting installations with linear sources of light, Philips Techn. Rev. **6**, 147-152, 1941.

3) K. Norden, Beleuchtungsberechnungen für Quecksilberdampflampen, Elektrotechn. Z. **28**, 757-758, 1907, and **29**, 883-886, 1908. See also E. L. Matthews, Das Licht **1**, 141-146 and 165-168, 1931.

4) See J. Loeb, Méthode générale de calcul de la répartition des éclairements dans les installations d'éclairage par lampes à fluorescence et autres sources linéaires, edited by Lampes Philips S.A., Geneva, 1948.

calculate what illumination an (infinitely small) element dl of the lamp produces in a point P of the work plane. The extremities of the lamp are denoted by *1* and *2* and their projections on the work plane by *1'* and *2'*. Further, the distance PP' from P to the line $1'2' = d$, and $P'1' = l_1$, $P'2' = l_2$. The radius vector from P to dl, of the length r, makes an angle β with the line PQ parallel to the lamp. Finally, arc tan $d/h = \alpha$.

If the luminous intensity of the element of the lamp (in a direction perpendicular to its axis) is $I_0 dl$ then the illumination at P perpendicular to the radius vector is $I_0 dl \sin \beta / r^2$. The horizontal illumination corresponding to this is

$$dE = \frac{I_0 dl \sin\beta}{r^2} \cdot \frac{h}{r} \quad \ldots\ldots \quad (1)$$

For a linear source of light the value of I_0 (the luminous intensity per unit length in a plane perpendicular to the lamp, in cd/m) can be derived from the luminous flux quoted by the manufacturers. This amounts to $\Phi/\pi^2 l \approx 0.1\ \Phi/l$, where Φ represents the luminous flux in lumens and l the length in metres. It must be borne in mind, however, that the definition of luminous intensity used in illumination engineering is based upon point sources of light, the luminous intensity of a lamp in a certain direction being determined by the luminous flux radiated in a small solid angle about that direction divided by that angle. In the case of linear sources of light this definition applies only when the observations are taken at a distance that is large compared with the dimensions of the lamp.

In the case of a point source of light the illumination (in lux) in a given point can be determined with the aid of a luxmeter and, if the distance from the lamp is known, from this it is possible to find the luminous intensity (in candelas). When the lamp is linear one can only deduce from the reading of a lux meter an "equivalent luminous intensity" of the source of light.

The factors dl and r in the second half of formula (1) will now be expressed in terms of h, α and β. Obviously $dl = r\,d\beta/\sin\beta$, and considering that $r = \sqrt{d^2 + h^2}/\sin\beta$ and $\sqrt{d^2 + h^2} = h/\cos\alpha$, we find that $r = h/\cos\alpha \sin\beta$. Substituting these expressions for dl and r in equation (1) the latter is transformed into:

$$dE = \frac{I_0 h \cos^2\alpha \sin^2\beta\, d\beta}{h^2} = \frac{I_0}{h} \cdot \cos^2\alpha \sin^2\beta\, d\beta. \quad (2)$$

The total illumination at P is found by integrating (2) between the limits β_1 and β_2, where $\beta_1 = \sphericalangle QP1$ and $\beta_2 = \sphericalangle QP2$.

We can therefore distinguish two cases. First we assume that the plane through P perpendicular to the lamp intersects the latter (in a point P'') as indicated in fig. 1. In this case, we shall first calculate the part $\int_{\beta_1}^{\pi/2}$ of the integral $\int_{\beta_1}^{\beta_2}$. This gives the illumination at P insofar as it comes from the part *1-P''* (length l_1) of the lamp, for which we find:

$$E_1 = \frac{I_0}{h} \cos^2\alpha \int_{\beta_1}^{\pi/2} \sin^2\beta\, d\beta =$$

$$= \frac{I_0}{h} \cos^2\alpha \left(\frac{\pi}{4} - \frac{1}{2}\beta_1 + \frac{1}{4}\sin 2\beta_1\right). \quad (3)$$

The position of the point P is determined by the distance d and the length l_1. Since $\cos\alpha = 1/\sqrt{1 + d^2/h^2}$ and $\tan\beta_1 = \sqrt{1 + d^2/h^2}\,.\,(h/l_1)$, from formula (3) E_1 can be calculated if d, l_1 and h are known. The factors (f) by which I_0/h has to be multiplied to arrive at the value of E_1 can be calculated and tabulated once for all for a number of values of d/h and l_1/h.

So far we have considered only a part of the lamp. The illumination E_2 given at P by the part *2-P''* (length l_2) of the lamp can, however, be calculated in exactly the same way. When d/h and l_2/h are known we can find this result from the same table of values as mentioned above. The total illumination E at P is then equal to the sum of E_1 and E_2.

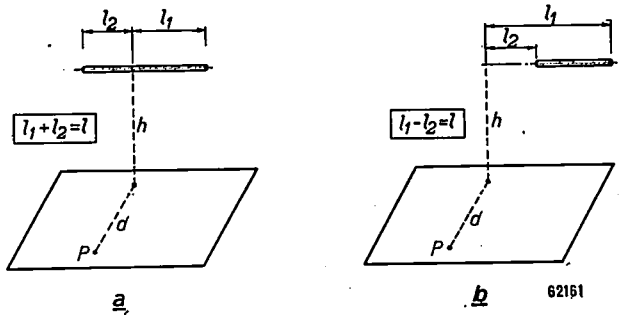
Secondly, we shall consider the case where the plane through P, perpendicular to the axis of the lamp, intersects the latter at a point on the extension of the axis (P_1'' in fig. 1). For l_1 we now have to take the distance from P_1'' to *1* and for l_2 that from P_1'' to *2*. Denoting again the illuminations from the hypothetical lamp parts P_1''-*1* and P_2''-*2* by E_1 and E_2 respectively, then E_1—E_2 gives the total illumination at P_1. (When P_1'' in fig. 1 lies on the other side of the lamp then it has to be regarded as the difference between the fictitious lamp parts P_1''-*2* and P_1''-*1*.)

Table 1 is the practical development of this discussion. It makes it quite easy to calculate the illumination in a point of the work plane whose position is determined by the distances d and l_1 (or l_2) when h has been measured and I_0 is known. If the light comes from a number of lamps in one or more planes parallel to the work plane then the total illumination in a point of the work plane is found by adding the illumination obtained from each lamp separately.

Lamp with fixture

To get a general picture of the distribution of the illumination on a horizontal measuring plane it is obvious that we must imagine a series of planes perpendicular to the axis of the lamp and then see how the illumination varies along the lines of intersection of these perpendicular planes with

Table I. The factors f for calculating the horizontal illumination E in a point P of the work plane.



For case (a):

$$E = I_0/h\,[f(d/h, l_1/h) + f(d/h, l_2/h)]$$

For case (b):

$$E = I_0/h\,[f(d/h, l_1/h) - f(d/h, l_2/h)].$$

l_1/h, l_2/h \ d/h	0	0.25	0.50	0.75	1.0	1.5	2	3	5
0.10	0.10	0.09	0.07	0.05	0.03	0.02	**0.01**	—	—
0.20	0.20	0.18	0.14	0.10	0.08	0.03	**0.02**	0.01	—
0.30	0.28	0.26	0.20	0.15	0.10	0.05	**0.03**	0.01	—
0.40	0.36	0.33	0.26	0.19	0.13	0.07	**0.03**	0.01	—
0.50	0.43	0.40	0.32	0.23	0.16	0.08	**0.04**	0.02	—
0.75	0.56	0.52	0.42	0.31	0.23	0.12	**0.06**	0.02	0.01
1.0	0.64	0.60	0.49	0.37	0.27	0.14	**0.08**	0.03	0.01
1.5	0.72	0.68	0.56	0.44	0.33	0.18	**0.11**	0.04	0.01
2	0.75	0.71	0.59	0.47	0.36	0.21	**0.12**	0.05	0.01
5	0.78	0.74	0.63	0.50	0.39	0.24	**0.15**	0.07	0.02
∞	0.79	0.74	0.63	0.50	0.39	0.24	**0.16**	0.08	0.03

the measuring plane. For such a perpendicular plane both l_1 and l_2 are constant. The factors by which I_0/h has to be multiplied for the successive points of the lines are found by taking the figures from one of the horizontal lines in table I (for the given l_1/h) and adding or deducting the corresponding figures in a second horizontal line (those for l_2/h).

We now suppose that the lamp is provided with a cylindrical fixture of the same length as the lamp and so narrow that the whole may be regarded as linear, and we shall assume that the radiation from the lamp and the fixture follows Lambert's law. Owing to the presence of the fixture the I_0 first used has to be multiplied by a certain factor depending upon d/h.

In order to find this factor we imagine that, from a point in a flat plane perpendicular to the lamp, vectors are plotted in each direction proportional to the I_0 of that direction. The line connecting the extremities of these vectors is then, for this plane, the distribution curve of the transversal luminous intensities I_0 depending upon the angle α [5]). In the case of a lamp without fixture this curve is a circle, for I_0 is then independent of α. Where the light source is in a fixture this distribution curve may assume many sorts of shapes. An example is given in *fig. 2*. The transverse luminous intensity $I_{0,\alpha}$ in the direction α is in this case represented by the product nI_0, where I_0 is the transversal luminous intensity of one lamp without fixture.

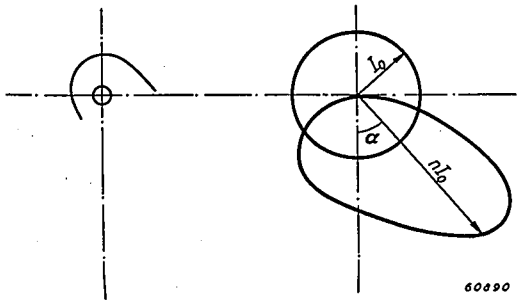


Fig. 2. On the right, distribution curve of the transverse luminous intensities in the case of a tubular lamp without fixture (circle with radius I_0), and the corresponding curve for an asymmetric fixture represented in cross section on the left.

In the curve representing the factor n of a fixture equipped with several parallel rows of lamps this number of rows is taken into account. The graph therefore contains a circle with radius $n = 1$ applying to one single row of lamps, whereas the vectors of the curve representing the factors n of the fixture relate to the totality of rows concerned.

It goes without saying that the values of the factor n can also be tabulated instead of plotting it in a graph. If the fixture is asymmetrical then on either side of the projection $1'\ 2'$ of the lamp on the horizontal plane in question the variation of n differs. In the case of the example drawn in fig. 2 we find the values:

d/h	3	2	1	0	1	2	3
α	71°30′	63°30′	45°	0°	45°	63°30′	71°30′
n	0	0.4	1.3	1.8	3,1	2.5	1.5

To find the illumination on the work plane for a point in the direction a, the illumination found

[5]) The distribution curve described here is not to be confused with the polar diagram of the light distribution of a point source of light in a meridional plane.

for a lamp without fixture has to be multiplied by the factor n corresponding to that direction.

A numerical example

We shall conclude this article by applying the method described to a simple example.

In a locality where the ceiling and walls are non-reflecting there are two parallel sources of light I and II mounted along the ceiling. The first source consists of an unbroken series of TL 25 W/33a fluorescent lamps freely suspended, whilst the second source is formed by two short TL 25 W/33a lamps with an asymmetric reflector, the light-distribution curve for which has been drawn in fig. 2.

It is required to determine the illumination along a line KL on a horizontal plane HH two metres below the light sources. (For the dimensions of the lamps and their position with respect to the line KL reference is made to *fig. 3*.)

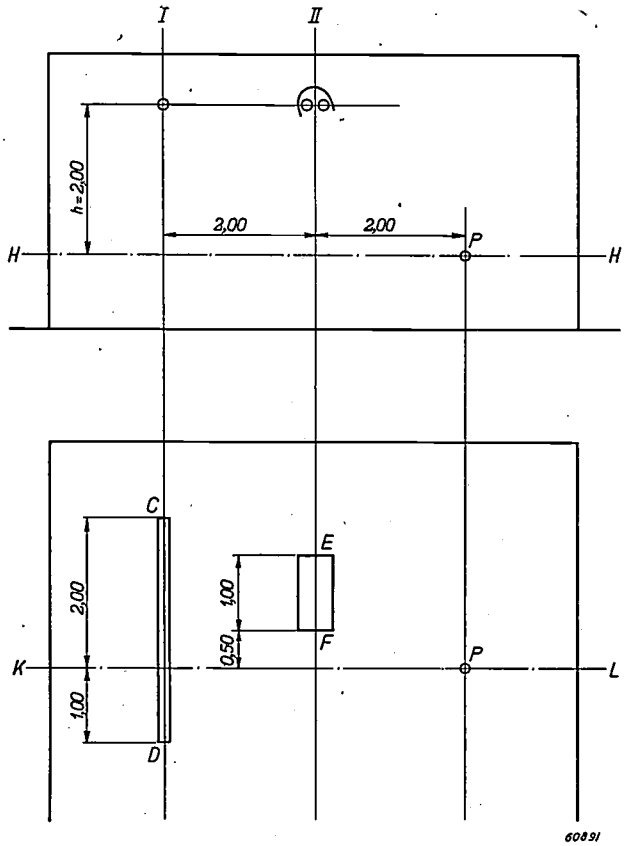


Fig. 3. Positional drawing indicating the dimensions of the two light sources I and II (referred to in the numerical example dealt with) and also their position with respect to the line KL in the work plane HH. It was required to calculate the distribution of the illumination along the line KL.

For both lamps $I_0 = 115$ cd/m, so that $I_0/h = 57.5$ cd/m².

To determine the illumination yielded by the lamp I we have to take for all points of KL in table I $l_C/h = 1.$ and $l_D/h = 0.5$, so that for the corresponding f_C and f_D we find:

d/h	3	2	1	0	1	2	3
f_C	0.03	0.08	0.27	0.64	0.27	0.08	0.03
f_D	0.02	0.04	0.16	0.43	0.16	0.04	0.02
f_C+f_D	0.05	0.12	0.43	1.07	0.43	0.12	0.05

In the same way the figures f_E and f_F for the source II are found by taking $l_E/h = 0.75$ and $l_F/h = 0.25$, thus:

d/h	3	2	1	0	1	2	3
f_E	0.02	0.06	0.23	0.56	0.23	0.06	0.02
f_F	0.01	0.02	0.09	0.24	0.09	0.02	0.01
f_E-f_F	0.01	0.04	0.14	0.32	0.14	0.04	0.01
n	0	0.4	1.3	1.8	3.1	2.5	1.5

The factors f_E-f_F then have to be multiplied by the coefficient n corresponding to each value of d/h.

The illumination in any point is then found by adding up the illuminations yielded by the light sources I and II. Let us take as example the point P (fig. 3) on the line KL. There the illumination amounts to

$$(0.12 + 3.1 \times 0.14) \times 57.5 = 32 \text{ lux.}$$

Summary. A simple method is discussed enabling one to calculate the illumination obtained in a certain point from linear sources of light. The calculation is restricted to the distribution of the illumination in a horizontal plane derived from direct lighting by means of one or more linear sources of light situated in a higher horizontal plane, for instance along the ceiling. The results of the calculations are compiled in a table. It is also explained how account can be taken of a fixture. Finally the theory is applied to a numerical example.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1899: G. Diemer and J. L. H. Jonker: Low-distortion power valves (Wireless Engr. **26**, 385-390, Dec. 1949).

A survey is given of various low-distortion valve constructions. Two new constructions are described by means of which the second harmonic of a single-stage pentode amplifier can be considerably reduced, resulting in a reduction of the total distortion by a factor of 2 up to an output of about 25% of the static anode dissipation. For even larger outputs (up to 50% of the static anode dissipation) the distortion is the same as that of a normal valve. The new valves have an I_a-V_a characteristic that is practically linear in the neighbourhood of the normal working point.

1900: L. A. Ae. Sluyterman and H. J. Veenendaal: A method for the detection of amphoteric substances in paper chromatography (Rec. Trav. Chim. Pays-Bas **68**, 717-720, 1949, No. 9/10).

Paper chromatograms are treated with a solution of tropaeolin OO, dried and held in hydrochloric acid vapour. Amphoteric substances appear as yellow spots on a red background.

1901: M. Asscher: A new synthesis of ω-amino-p-hydroxyacetophenones and their reduction to the corresponding amino-aethanols (Rec. Trav. Chim. Pays-Bas **68**, 960-968, 1949, No. 11).

A new synthesis is described of compounds of the type

$$HO-\langle\overline{=}\rangle-CO\,.\,CH_2N\langle$$

by the condensation of phenol or its derivatives with aminoacetonitriles according to the method of Hoesch and Houben. The corresponding alcohols are prepared by catalytic reduction of the ketones obtained.

1902: F. A. Kröger, J. E. Hellingman and N. W. Smit: The fluorescence of zinc sulphide activated with copper (Physica The Hague **15**, 990-1018, 1949, No. 11/12).

The fluorescence of zinc sulphide activated with copper consists of three bands: a green band and a blue band caused by copper, and the blue band of self-activated zinc sulphide. In addition to fluorescence centres copper may also give rise to quencher centres. The fluorescence centres have an absorption band associated with them, situated at the long-wave end of the fundamental absorption of zinc sulphide.

The variation in the relative concentration of the various centres with the conditions of preparation has been studied, using controlled atmospheres of H_2S-HCl, H_2-HCl, etc. At high temperatures ($T > 1050$ °C) high HCl contents in the atmosphere enhance the formation of green centres while low HCl contents favour the formation of blue copper centres, but neither of these centres are formed when HCl is entirely lacking. High concentrations of H_2S favour the formation of quencher centres. When products made at high temperatures are fired at a lower temperature (e.g. 400 °C), reducing atmospheres like H_2, N_2 (and also O_2!) enhance the blue centers and destroy quencher centres, while oxidizing atmospheres (H_2S, HCl) have the opposite effect. These effects can be explained if it is assumed that the green copper centre is formed by CuCl, the blue copper centre by CuCl-Cu, the blue zinc centre by ZnCl and the quencher centre by Cu_2S, all being incorporated in the zinc sulphide lattice. The physical behaviour of the system is discussed on the basis of the Schön-Klasens theory of energy transfer between centres.

1903: C. J. Bouwkamp: On the evaluation of certain integrals occurring in the theory of the freely vibrating circular disk and related problems (Proc. Kon. Ned. Akad. Wetenschappen Amsterdam **52**, 987-994, 1949, No. 9; Indigationes Mathematicae **11**, 366-372, 1949, fax. 5).

In the theory of the acoustic radiation by a freely vibrating rigid circular disc integrals of the form

$$\int_0^1 P_{2n+1}(\sqrt{1-\varrho'^2})\varrho'\,d\varrho' \int_0^{2\pi} [\varrho^2-2\varrho\varrho'\cos\vartheta'+\varrho'^2]^{\frac{1}{2}(m-1)}\,d\vartheta'$$

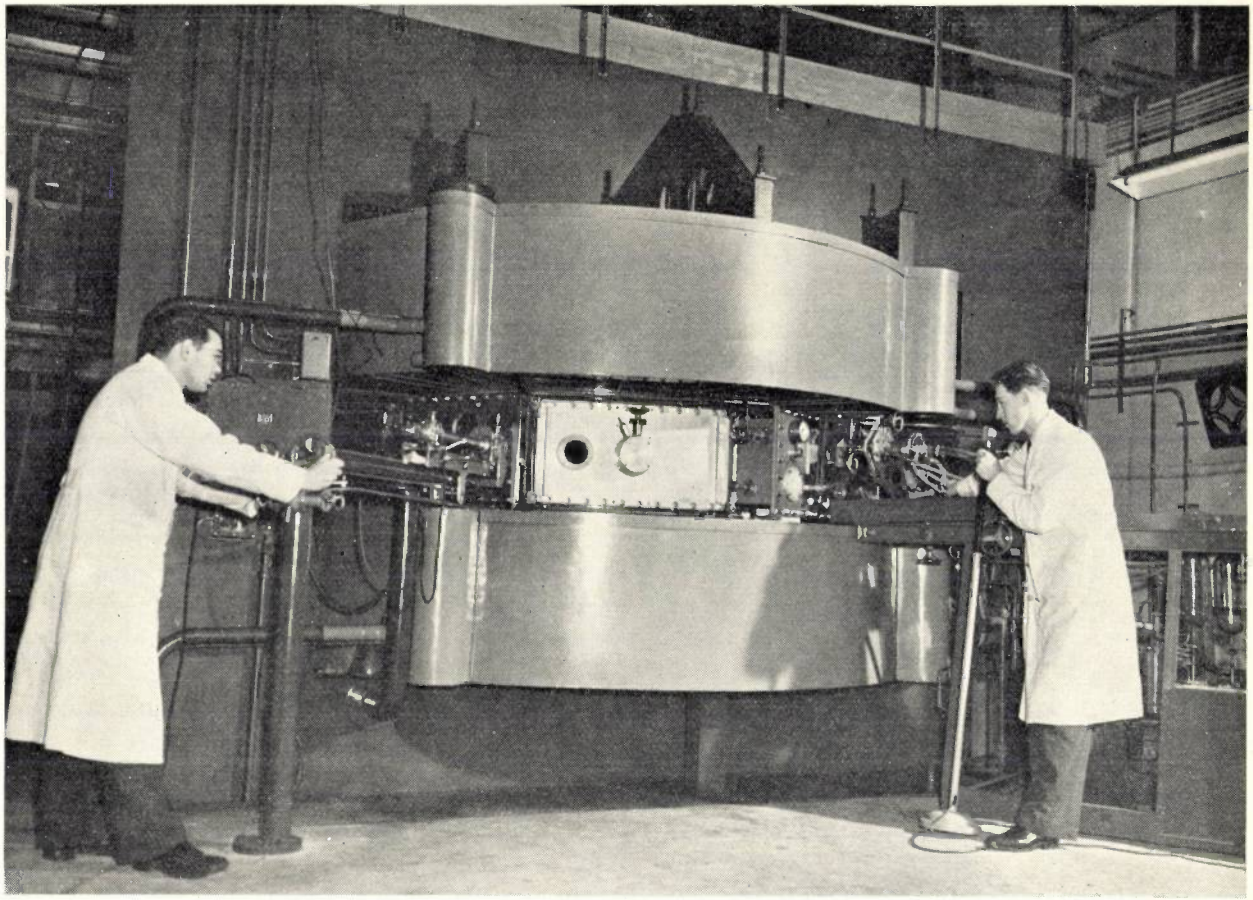
occur in which P is a Legendre polynomial. The same integrals occur in the theory of acoustic diffraction by a circular disc or by a circular aperture. It appears that these integrals are polynominals in ϱ^2. These may be written either in the form of a hypergeometric series or as a linear combination of Legendre polynomials.

VOL. 12 No. 3, pp. 65-96 SEPTEMBER 1950

Philips Technical Review

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

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



61995

CYCLOTRON AND SYNCHROCYCLOTRON

by W. de GROOT.

621.384.61

On 10th November 1949, in the Institute for Nuclear-Physical Research at Amsterdam, a synchrocyclotron built by Philips for the said Institute to a design by Prof. Dr. C. J. Bakker and Prof. Dr. Ir. F. A. Heyn was officially taken into use. It is intended to publish a series of articles in this journal describing the construction and working of this apparatus and giving some information about the research work that can be done with it, as also about its further possibilities.

By way of introduction a more general account is given of the working of the cyclotron and the synchrocyclotron, mainly taken from literature already published on this subject.

Introduction

To impart a high velocity and thus a great kinetic energy to a small mass which can travel along a prescribed rectilinear or curved path without friction, a force has to be applied to that mass. The increase ΔT in the kinetic energy T along a part of the path is given by

$$\Delta T = \int K_s ds, \quad \ldots \ldots \quad (1)$$

where K_s is the component of the force along the path and the integration has to be carried out over the part of the path in question. Any component of the force perpendicular to the path and the normal reaction due to the path (which force is likewise perpendicular to the path in the case of frictionless movement) do not perform any work and thus do not contribute towards the increase of T.

Instead of (1) it is sometimes advantageous to use the formula for the increase of momentum (p), which reads:

$$\Delta p = \int K_s dt \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In classical mechanics we have:

$$T = \tfrac{1}{2} M_0 v^2 , \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$p = M_0 v , \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where M_0 represents the mass of the particle.

Since, however, we shall presently have to deal with particles having a velocity approaching that of light, we shall also have to take into account the changes taking place in the expressions for T and p when v approaches the velocity of light c and thus the quotient v/c, which is always less than unity, becomes of the order of unity. Formulae (1) and (2) then remain valid, but for T and p we get:

$$T = Mc^2 - M_0 c^2 , \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$p = Mv , \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where M_0 now represents the mass of the body at rest and

$$M = \frac{M_0}{\sqrt{1 - v^2/c^2}} . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

It is obvious that for $v \ll c$ (5) and (6) become identical with (3) and (4).

The force K may in various ways depend upon the time t or the place s on the path. Let us consider a few cases:

1) An example of the case where K is constant and the particle describes a straight path is, for instance, when a stone is dropped from a high tower.

2) In the case of a projectile shot out of a gun — disregarding the force of gravity — we are concerned with a force K which is very great for a short time and after that zero. Thus the body is given the desired energy in one impulse.

3) One can also imagine the particle being accelerated by a series of impulses at regular distances.

4) In the case of a particle of mass being propelled along a straight path by periodic impulses of short duration, owing to the increasing velocity the distance travelled in the intervals becomes greater and greater; that is to say, the distances measured along the path from point to point where the pulse begins to take effect become longer and longer.

5) All these modes of acceleration can be realized when the particle is caused to describe a curved instead of a straight path for instance by forcing it to travel without friction through a curved channel or like a frictionless bead along a curved wire.

The acceleration of atomic particles

In nuclear physics use is made of light nuclei (protons, deuterons, alpha particles) for bringing about nuclear reactions [1]). Since the nuclear particles acting as projectiles are electrically charged, it is a fairly simple matter to bring accelerating forces to act upon them by subjecting them to the action of electric fields. All cases occurring in the mechanical examples given above are in fact realized in practice.

The acceleration of protons and deuterons by an electric field constant in space and time (analogous to the falling stone) takes place in ion-accelerating tubes for not too high tensions (e.g. 50 to 100 kV).

Acceleration by one short impulse is realized, for example, in the cathode ray tube, in which electrons are accelerated between a cathode and an anode placed close to it, together forming what is called an electron gun, the electrons then travelling on at a uniform velocity. This method of acceleration is also applied to ions, as for instance in the mass spectrograph.

Similar to this is the action taking place in the atoms of a radioactive substance emitting alpha particles. Upon emerging from the nucleus the alpha particle receives an impulse from the repelling Coulomb force of the nucleus and then continues on its way at a uniform velocity until ultimately, retarded by collisions with the molecules of the air, it reaches the end of its track.

As an example of acceleration by short pulses applied at regular distances along the path we may take the action of ion-accelerating tubes for tensions of some hundreds of kilovolts, where for technical reasons the tube is divided into a number of sections, a fraction of the total potential difference being applied to each section.

Acceleration by periodic pulses along a rectilinear

---

[1]) For a general treatise on fundamental nuclear reactions and the part played therein by the particles mentioned, see W. de Groot, Philips Techn. Rev. 2, 97-102, 1937.

[2]) See, e.g., F. A. Heyn and A. Bouwers, An apparatus for the transmutation of atomic nuclei, Philips Techn. Rev. 6, 46-53, 1941.

path takes place in what is called the linear accelerator, an apparatus which it is not necessary to discuss here.

Finally, acceleration by periodic pulses along a curved path takes place in the cyclotron constructed by Lawrence [3]), the apparatus to which this present article is devoted. The "normal force" required to cause the particle to follow an orbit is derived, in this cyclotron, from a magnetic field: when a particle with mass M and charge Q travels at a velocity v in a direction perpendicular to the lines of force of a magnetic field with induction B it experiences a force K (Lorentz force) perpendicular to v and B given by

$$K = Q\,v\,B\,.$$

The radius of curvature r of the orbit is thereby given, since this force must be equal to the centripetal force required, so that

$$\frac{Mv^2}{r} = Q\,v\,B$$

or

$$Q\,B\,r = M\,v = p\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If v is comparable to the velocity of light c then M is again given by equation (7).

If B is constant then, with a given velocity v, by (8) we have a circular path, since M is determined only by v. The particles then travel along this circle with an angular velocity

$$\omega = \frac{v}{r} = \frac{QB}{M}\,. \quad . \quad . \quad . \quad . \quad . \quad (8a)$$

When the particle is accelerated, so that v increases with time, then also r increases and the path described becomes a spiral.

The cyclotron

The cyclotron consists in the main of a magnetic circuit with an "air gap" between the poles (see *fig. 1*). This gap may be of rather considerable dimensions. The distance between the poles is a few decimetres and the cross-sectional diameter of the poles is usually more than 1 metre. Between the poles, which are placed vertically one above the other, forming an airtight joint with them, is a ring of non-magnetic material isolating the gap from the outside air. The cylindrical box-shaped space thus formed is evacuated. Inside this space and supported by insulated rods is a smaller cylindrical box bisected diametrically into two halves each shaped like the capital letter D. That is why these two halves are commonly referred to as the "dees". The gap between the dees is a few cm wide.

Between the dees, which together form a sort of capacitor, an alternating voltage of some tens of kV is applied with a frequency in the order

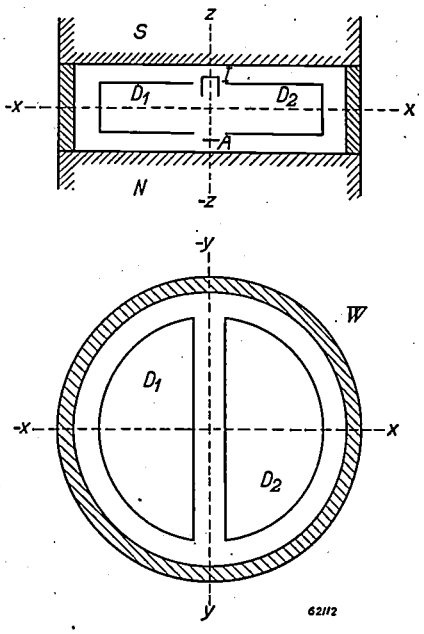


Fig. 1. The acceleration chamber of the cyclotron (drawn schematically). N and S the two poles of the magnet, W the wall of the vacuum chamber, D_1D_2 the two "dees", I ion source (in which is a filament), A anode for picking up the electrons coming from the source.

of 10^7 c/s. As a result an approximately horizontal, homogeneous, electric alternating field is formed in the gap between the dees, while the remaining space inside the dees is practically free of any field. In addition, everywhere in the gap there is a constant and practically homogeneous magnetic field with induction B of about 1.5 Wb/m^2. In the centre of the gap, slightly above the horizontal plane of symmetry, is an ion source by means of which deuterons, for instance, can be injected into the vacuum at a low velocity.

Let the electric field between the dees be

$$E = E_{\mathrm{m}} \cos \omega_E t\,,$$

with ω_E given a very definite value such, as will presently be explained, as to equal the aforementioned angular velocity of the particles, i.e., for not too high velocities, where $M = M_0$, a value:

$$\omega_E = \frac{QB}{M_0}.$$

Let us consider a particle at rest in the field. This will be given momentum by the field and as soon

[3]) The first publication on this subject was by E. O. Lawrence and N. E. Edlefsen, Science 72, 376-377, 1930.

as it reaches a certain velocity the Lorentz force comes into action, thereby causing the particle to describe a sort of spiral path lying for the moment in the gap between the dees. The shape of this path depends to some extent upon the initial conditions, in particular upon the moment at which the particle (with velocity zero) is released into the field.

After a few loops, however, a situation arises where the particle lies approximately in the middle of the gap at the moment that the field is at its maximum and where the radius vector r of the orbit increases proportionately with t.

The differential equations for a particle with charge Q and mass M_0 in a horizontal electric alternating field

$$E_x = E_m \cos \omega_E t ,$$

combined with a vertical, homogeneous, magnetic field B_z, where

$$\omega_E = \frac{Q}{M_0} B_z ,$$

are, with the coordinates indicated in fig. 1:

$$M_0 \ddot{x} = Q(E_x - B_z \dot{y}) ,$$
$$M_0 \ddot{y} = QB_z \dot{x} ,$$

or, omitting the index E,

$$\ddot{x} = (QE_m/M_0) \cos \omega t - \omega \dot{y} ,$$
$$\ddot{y} = \omega \dot{x} .$$

The solution for a particle released into the field at a moment $t = \tau$ (with $x = y = 0$ and $\dot{x} = \dot{y} = 0$ for $t = \tau$) is:

$$x = \frac{QE_m}{2M_0\omega^2} [-\sin \omega\tau \sin \omega (t-\tau) + \omega (t-\tau) \sin \omega t] ,$$

$$y = \frac{QE_m}{2M_0\omega^2} [-\cos \omega\tau \sin \omega (t-\tau) + 2 \sin \omega\tau (1 - \cos \omega (t-\tau)) + \omega (t-\tau) \cos \omega t] .$$

These equations represent a sort of spiral the shape of which depends to a certain extent upon the parameter τ. After the particle has travelled one single loop $\omega(t-\tau) \gg 1$ and it is mainly the last terms that are of importance. If, for instance, $\omega(t-\tau) = 2\pi$, $E_m = 10^6$ V/m and $B_z = 1.5$ Wb/m^2, then for a deuteron ($\omega = 7 \times 10^7$ rad/sec) the radius vector $r = (x^2 + y^2)^{1/2}$ is equal to $\pi QE_m/M_0\omega^2 = \pi E_m/\omega B_z \approx 3$ cm.

Owing to the increase of r the particle soon penetrates into the field-free space inside the dees. From that moment onwards it is accelerated only during the periods of time when it traverses the gap. In the interim period it describes semi-circular paths with a constant velocity, with a slightly larger radius each time.

The time taken to describe such a semi-circle is: $\pi r/v = \pi M_0/QB = \pi/\omega_E$. From the value chosen for ω_E it follows that once the particle is made to traverse the gap at the moment that the field is at its maximum it will continue to do so even though the radius of the orbit increases each time; to put it in other words, the particle is in resonance with the electric field.

Each time the particle traverses the gap its energy increases by a constant amount ΔT. According to formula (3) therefore v^2 is proportional to the number of loops and, since v is proportional to r, the radius of the spiral increases further in proportion to $\sqrt{t}$, so that the wider the turns of the spiral the closer they will lie together.

The greatest value that the radius can assume is determined by the radius R of the dees. When this is reached then, for an ion with charge $Q = Ze$ — where e represents the elementary charge (= 1.6×10^{-19} coulomb) and Z the charge number of the ion (for light ions equal to the nuclear charge) — the energy is:

$$T_{max} = \tfrac{1}{2} M_0 v^2 = \tfrac{1}{2} M_0 \left(\frac{RQB}{M_0}\right)^2 = \frac{Z^2e^2}{M_0} \cdot \frac{R^2B^2}{2} .$$

Since

$$M_0 \approx A\, M_H ,$$

where M_H represents the proton mass and A the atomic weight, and expressing the energy in electron volts, it is found that with $T_{max} = eV_{max}$:

$$V_{max} = \frac{Z^2}{A} \cdot \frac{e}{M_H} \cdot \frac{R^2B^2}{2} .$$

If, for instance, $R = 0.5$ m and $B = 1.5$ Wb/m^2 then, since $e/M_H \approx 10^8$ coulombs/kg, we have:

$$V_{max} \approx 3 \cdot 10^7 \frac{Z^2}{A} \text{ volts} = 30 \frac{Z^2}{A} \text{ megavolts.}$$

For protons $Z = 1$ and $A = 1$, whilst for deuterons $Z = 1$, $A = 2$ and for alpha particles $Z = 2$, $A = 4$. Thus in the case considered here the maximum energy for deuterons is about 15 MeV and for protons and alpha particles 30 MeV.

These figures, however, are only of value for the order of magnitude, because, as will presently be shown, there are other factors affecting the maximum energy attainable.

Electric and magnetic focusing

It has been assumed above that the electric field between the dees is homogeneous and that the electric force is parallel to the horizontal plane of symmetry of the system. A particle released with zero velocity will then begin to move in a direction which is also parallel to that horizontal plane. This, however, is all only very approximately true. Actually the electric equipotential planes penetrate somewhat into the space inside the dees and the

electric lines of force are as a result curved. *Fig. 2* illustrates this effect. One of the consequences of this is that those particles which do not start exactly in the central plane receive momentum in the vertical direction and soon reach the top or bottom walls of the dees.

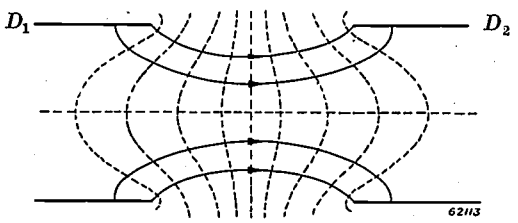


Fig. 2. The electric field between the dees.

The inhomogeneity of the electric field is of importance also during the further process of acceleration, but on the whole this influence is favourable, since the field appears to have a concentrating action, drawing back to the horizontal plane of symmetry such particles as, through some cause or other, have received a velocity component leading them away from that plane [4]) [5]). This is due to two causes. In the first place, owing to their velocity increasing in the field, the particles are subjected to diverging forces (on the right in fig. 2 if the particle comes from the left) for a shorter time than they are exposed to converging forces (on the left in fig. 2), and moreover many particles pass through the field while it is already beginning to decrease, thus assisting the converging action.

Since the influence of the field upon the electrons is comparable to that of a lens upon a beam of light rays one speaks of electric "focusing". This focusing action diminishes as the velocity of the particle increases, ceasing when $r \approx {}^1/_3 R$.

Fortunately a second focusing then comes into action, namely the magnetic focusing. Owing to the spread of the magnetic field there is a small decrease in the strength of that field from the centre outwards as a result of which the lines of force curve slightly outwards (*fig. 3*) and consequently the Lorentz force at points outside the centre plane has a component perpendicular to and directed towards that plane. Thus the particle is subjected to a force proportional to the distance from the plane of symmetry and begins to oscillate about the ideal path situated in that plane with a frequency which, as a further calculation shows, is given by:

$$\omega_{\text{vert}} = \omega_E \sqrt{n},$$

where $n = -\mathrm{d} \log_e B_z / \mathrm{d} \log_e r$.

[4]) M. E. Rose, Phys. Rev. **53**, 392-408, 1938.
[5]) R. R. Wilson, Phys. Rev. **53**, 408-420, 1938.

There is also a horizontal magnetic focusing action: particles which, for some reason or other (e.g. owing to collisions with gas molecules), have either a direction of velocity or a radius vector differing from that of the ideal path are automatically attracted towards the ideal path again. This, too, is accompanied by oscillations, with a frequency

$$\omega_{\text{hor}} = \omega_E \sqrt{1-n}.$$

The existence of "vertical stability" can be proved as follows [6]). If z is the distance from the horizontal plane of symmetry then

$$M_0 \ddot{z} = Q v B_r,$$

where B_r represents the radial component of the induction.

In the gap, according to Maxwell's equations:

$$\frac{\partial B_z}{\partial r} - \frac{\partial B_r}{\partial z} = 0.$$

Now in the plane of symmetry $B_r = 0$ and $z = 0$. Hence $\partial B_r/\partial z$ is to a first approximation equal to B_r/z, so that

$$B_r = z \frac{\partial B_z}{\partial r}.$$

This gives:

$$\ddot{z} = \frac{Q\omega}{M_0 r} z \frac{\partial B_z}{\partial r},$$

and since $QB/M_0 = \omega \approx \omega_E$:

$$\ddot{z} + \omega_E^2 z \frac{r}{B_z} \frac{\partial B_z}{\partial r} = 0,$$

or:

$$\ddot{z} + \omega_E^2 n z = 0.$$

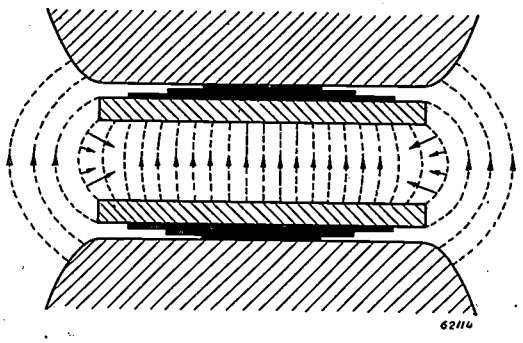


Fig. 3. The magnetic field of the cyclotron with lines of force curved outwards (dotted lines). The strength of the field diminishes slightly from the centre outwards. The direction of the Lorentz force at the edge is indicated by small arrows.

This represents an oscillation with the frequency $\omega_E \sqrt{n}$. In a similar way the existence can be proved of radial stability with the frequency $\omega_E \sqrt{1-n}$.

If $n = 0.2$ then $\omega_{\text{hor}} = 2\omega_{\text{vert}}$. Closer investigation shows that a sort of resonance then arises, accompanied by transmission of energy between the two modes of oscillation and resulting in considerable widening of the path and consequent defocusing. That is why care has to be taken to ensure that

[6]) D. W. Kerst and R. Serber, Phys. Rev. **60**, 53-58, 1941. See also A. Bierman and H. A. Oele, Betatrons with and without iron yoke, Philips Techn. Rev. **11**, 65-78, 1949, No. 3.

the fraction n, which has to be as small as possible with a view to keeping the particles "in step", does not exceed the value 0.2. The radial variation of B_z and thus the magnitude of n can be altered by providing an additional air gap and partly filling this with circular iron plates ("shims") of different diameter (see fig. 3). It can be so arranged, for instance, that in the centre B_z decreases with the square of r and farther away from the centre linearly with r. In this way the above condition for n can be satisfied in practically the whole of the accelerating space ($0 < r < R$).

Limitations of the classical cyclotron

According to the theory outlined above the energy of the particles could be increased indefinitely by increasing the radius R of the dees. Increasing B would have the same effect but then, of course, one is confined to material limits.

Actually there are other natural limits for the highest energy attainable. In the first place, for the magnetic focusing it is essential that the field diminishes outwards. In the second place, according to (5) the quantity M increases with increasing energy. As a result of these causes the quantity $\omega = QB/M$, i.e. the value the electric angular frequency should have for exact resonance of the particles circling with the electric field, diminishes as the velocity increases, thus as the radius of the orbit becomes larger.

Eliminating v from the equations (7) and (8a) we find ω and M as functions of r, viz.:

$$\frac{1}{\omega^2} = \left(\frac{M_0}{QB}\right)^2 + \frac{r^2}{c^2},$$

$$M^2 = M_0^2 + \left(\frac{QBr}{c}\right)^2.$$

If $\omega_{r=0} = \omega_E$ is the constant frequency of the electric field then at the edge of the gap ω_E will be greater than $\omega(r)$.

As a result, with a certain value of r the particle will travel rather too slowly and next time pass the gap a little too late. If this effect accumulates the particle will ultimately reach the gap at a moment when the electric field is zero or even of opposite polarity, so that it no longer has an accelerating action and may even decelerate the particle, the energy no longer increasing but even possibly decreasing.

This can be remedied by giving ω_E a value in between $\omega_{r=0}$ and $\omega_{r=R}$, and further by giving the electric field between the dees the largest possible amplitude, so that the acceleration each time the particle passes through the field is as great as possible and the number of loops that have to be travelled to reach a certain energy is as small as possible, while the particle "loses step" as little as possible. Calculations show that with an amplitude of some hundreds of kV between the dees it is possible in this way, with a reasonable yield, to accelerate protons up to 15 MeV, deuterons up to 25 MeV and alpha particles up to 50 MeV (see the article quoted in footnote [4]). This, however, is the limit for the "classical" cyclotron.

The synchrocyclotron

When, about 1930, E. O. Lawrence conceived the cyclotron (see the article quoted in footnote [3]), thereby basing the design on the resonance formula $\omega_E = QB/M_0$, and decided to construct such an apparatus, it was not a priori certain that he would succeed in getting particles of great energy with reasonable efficiency. However, fortune favours the bold, and fortune here was in the form of the electric and magnetic focusing, thanks to which a practical application of the idea first became possible.

It has been seen, however, that by the very reason of B_z decreasing with increasing r — which condition is essential for magnetic focusing — and moreover owing to M increasing with increasing T, it is impossible to fulfil the resonance condition exactly and that this sets a limit to the highest energy attainable.

Now, theoretically, there is a simple means of meeting this, namely by making the frequency ω_E of the electric field variable instead of leaving it constant, so that the condition $\omega_E = QB_z/M$ will always be valid while the particle is describing its spiral.

Again it is not a priori certain that this principle can be successfully applied with a high efficiency. It could be argued that this condition can only be fulfilled for one particle at a time, namely for the particle that starts off just at the right moment, and not for particles starting a little earlier or later. Furthermore, it seems difficult to "modulate" the frequency ω_E in such a way that the condition $\omega_E = QB_z/M$ is always exactly fulfilled along the whole of the orbit.

Here too, however, there is a fortunate circumstance which makes it possible to apply successfully the principle published almost simultaneously in 1945 by Veksler and by Mc Millan [7]), and which, moreover, relieves us of the necessity to make ω_E dependent upon time in a prescribed manner.

---

[7]) V. Veksler, J. Phys. U.S.S.R. 9, 153, 1945; E. M. Mc Millan, Phys. Rev. 68, 443L, 1945. — An analogous proposition was made by Oliphant in 1943, but not published until 1947: M. L. Oliphant, J. S. Gooden and G. S. Hide, Proc. Phys. Soc. 59, 666, 1947.

To make this clearly understood let us once more assume that ω_E is constant and that there is a radius vector r for which

$$\omega_E = \omega(r) = \frac{v(r)}{r} = \frac{QB(r)}{M} \quad . \quad . \quad . \quad (9)$$

Imagine that there is a particle travelling along this circular path and passing the gap between the dees each time at the moment at which the electric field is zero and changing from an accelerating to a decelerating field. Let us say that such a particle has a phase $\varphi = 0$ with respect to the field.

It is obvious that since its energy T does not change this particle can continue travelling along this orbit for an indefinite length of time. Now imagine a particle describing the same orbit but such that at the moment when the field is zero its azimuth with respect to the gap has a value φ_1. Each time this particle traverses the gap the field will have an accelerating action. Therefore, each time it passes the gap it will be accelerated, the energy T thereby increasing, so that automatically the radius r of the orbit is increased and the looping frequency reduced. Thus the particle will reach the gap a little later each time until it again has a phase $\varphi = 0$ with respect to the field. Then, however, it will no longer satisfy equation (9), because its ω is then smaller than ω_E and thus the particle still travels too slowly.

The process therefore continues in this way: in course of time the particle will come to pass the field at a moment when it has a retarding action, as a consequence of which its energy diminishes and thus r decreases and $\omega(r)$ again becomes greater. Ultimately it once more arrives in its old orbit with the right $\omega(=\omega_E)$ but now with a negative phase, exactly opposed to the phase φ_1 with which it started ($\varphi = -\varphi_1$). The process is then repeated in the reverse direction, the phase increases (thus decreasing in absolute value), ω increases, while T decreases. When, as a result of this, φ has again become zero we still find $\omega > \omega_E$. Thus the increase of φ continues further, accompanied by a decrease of ω and increase of T. When ultimately the particle has again arrived in its old orbit it once more has the phase φ_1 with which it started and the original values of ω and T. This cycle is repeated for an indefinite length of time and we thus see a sort of vibration set in whereby the phase oscillates between φ_1 and $-\varphi_1$ while at the same time T, ω and r fluctuate about a mean value.

Thus there is a certain stability (phase stability) with respect to the orbit determined by (9). This stability is maintained up to very large values of φ_1, namely up to $\varphi_1 = \pm\pi$.

Let us now imagine the same process taking place while ω_E is not constant but gradually diminishing.

A particle that is describing an orbit whereby $\omega = \omega_E$ but which passes the gap at the moment when the field is accelerating (thus φ is not zero but positive) will increase or decrease in phase as ω_E becomes smaller. This depends upon whether the decrease of ω, accompanied by the increase of T, is greater or less than the decrease of ω_E. One can therefore imagine that there is a particle with such a phase ($\varphi = \varphi_s$) that for this particle ω always equals ω_E. This particle will continually be accelerated and will always be synchronous with the field (it is from this synchronism that the name synchrocyclotron has been derived). It can, however, also be proved that particles whose phase at a certain moment is not equal to φ_s and for which at that moment ω is not equal to ω_E will start oscillating about the expanding "synchronous" orbit and thus on the average be accelerated just as much, provided $|\omega - \omega_E|$ and $|\varphi - \varphi_s|$ remain confined within certain limits.

From the foregoing it follows that when $\varphi = \varphi_s$ and $\omega = \omega_E$, then $d\varphi/dt = 0$. It is obvious that in general $d\varphi/dt = \omega - \omega_E$ and thus $d^2\varphi/dt^2 = d\omega/dt - d\omega_E/dt$, in which the differentiation process has to be carried out such that dt does not diminish to zero but remains great enough for at least one loop to be completed in the time dt. With this assumption the differential equation for φ reads [8]):

$$\frac{d^2\varphi}{dt^2} = C(\sin\varphi_s - \sin\varphi) \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where

$$C\sin\varphi = -d\omega/dt \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and

$$C\sin\varphi_s = -d\omega_E/dt \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Equation (10) is the same as that for a mathematical pendulum upon which, in addition to the usual "gravitation couple" $-C\sin\varphi$, a constant couple $C\sin\varphi_s$ is acting (see *fig. 4*). If $\varphi_s = 0$ ($\omega_E = $ const.) then φ fluctuates about the value $\varphi = 0$ between $+\varphi_1$ and $-\varphi_1$, with $|\varphi_1| < \pi$. If $\varphi_s \neq 0$ then the state of equilibrium is $\varphi = \varphi_s$ and φ may fluctuate between φ_1 and φ_2, where $\varphi_s \leqq \varphi_1 \leqq \pi - \varphi_s$, whilst $-(\pi - \alpha) \leqq \varphi_2 \leqq \varphi_s$. α increases with increasing $\varphi_s (\alpha > 2\varphi_s)$.

In the case of the cyclotron C and φ_s, strictly speaking, are functions of r, just like B, ω and T. Instead of (11) one can in fact also write:

$$C\sin\varphi = -\frac{d\omega}{dT}\frac{dT}{dt} = -\frac{d\omega}{dT}\cdot 2eV\sin\varphi\cdot\frac{\omega}{2\pi},$$

---

[8]) D. Bohm and L. Foldy, Phys. Rev. **70**, 249-258, 1946; **72**, 649-661, 1947. In these articles the energy gain per revolution is taken as being equal to $eV\sin\varphi$. In the present article $2eV\sin\varphi$ has been taken, where V represents the maximum potential difference between the dees and account has been taken of the fact that the particle traverses the gap twice in every revolution.

hence:

$$C = -\frac{d\omega}{dT} \cdot 2eV \cdot \frac{\omega}{2\pi}.$$

The variation of C with r however is not so great (10-20%) as to invalidate the truth of the following argument where C is regarded as a constant.

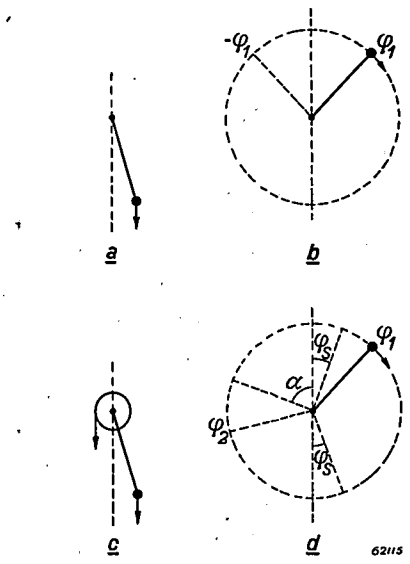


Fig. 4. Diagrammatic drawing of a mathematical pendulum behaving according to formula (10).
a) and *b*) $\varphi_s = 0$, the ordinary mathematical pendulum, extreme positions $+\varphi_1$ and $-\varphi_1$, whereby $|\varphi_1| < \pi$.
c) and *d*) pendulum with constant couple $C \sin \varphi_s$, extreme positions φ_1 and φ_2, whereby $\varphi_s \leqq \varphi \leqq \pi - \varphi_s$ and φ_2 lies between φ_s and $-(\pi - \alpha)$, $(\alpha > 2\varphi_s)$.

From (12) it follows that $-d\omega_E/dt$ may not be greater than C, as otherwise it is impossible to have a synchronous path. The more C differs from $-d\omega_E/dt$ the larger is the group of particles "taken along". But $-d\omega_E/dt$ may not be too small either, because then, owing to the phase oscillation, a large number of particles tend to return to the starting point and are lost as far as the acceleration is concerned. It must be borne in mind that, as already shown, almost all the particles begin with a phase $\varphi = \pi/2$, where $\dot{\varphi}$ may be either greater or less than 0 according to whether, at the moment that the particle starts off, $\omega_{r=0}$ is greater or less than the instantaneous value of ω_E. Thus there is an optimum value for $-d\omega_E/dt$, and at that value the portion of all particles carried along is as large as possible.

Actually the frequency ω_E is "modulated" by connecting to the oscillator supplying that frequency a variable capacitor the capacitance of which is a periodical function of time, with a cycle that is large compared with the time taken by a particle to complete one orbit.

As a function of t, therefore, ω_E increases and decreases alternately, and only those intervals of time in which ω_E decreases are useful for the acceleration. Due to the phase stability, however, not one particle but a whole group of particles benefits from the favourable conditions. In practice this group comprises, say, 1% of the total number of particles, so that 99% of the ions produced are lost for the acceleration. With the synchrocyclotron this adverse factor has to be taken into account. On the other hand there is the favourable factor of being able to increase the energy output.

Another advantage of the method described is that — since there is no fear of the particle getting out of phase with respect to the field — there is no imperative need to ensure that the number of "loops" required to obtain a certain energy is as small as possible. That is why, in the synchrocyclotron, the voltage between the dees need not be so high as in the case of the classical cyclotron, a factor that has a number of practical advantages in connection with insulation and the risk of sparking. In particular, less stringent requirements have to be met as regards the vacuum, thus allowing of a large production of ions in the centre.

Summary. Following upon an introduction on the acceleration of particles as a mechanical problem, a description is given of the functioning of a cyclotron in which light nuclei, such as protons, deuterons or alpha particles, can be accelerated up to an energy of some tens of millions of electron volts. A limit is set to the energy attainable by the fact that, owing to the magnetic induction decreasing and the mass increasing for velocities comparable to the velocity of light, at the edge of the field the resonance condition is no longer fulfilled. It is then indicated how, in the synchrocyclotron, by modulating the frequency of the accelerating field the resonance condition can, on an average, be fulfilled right up to the edge of the field, at least for a portion of the particles. As a result, be it at the cost of efficiency, particles can be obtained with very much higher energies than are possible with the classical cyclotron.

AN EXPERIMENTAL "STROBOSCOPIC" OSCILLOSCOPE FOR FREQUENCIES UP TO ABOUT 50 Mc/s

II. ELECTRICAL BUILD-UP

by J. M. L. JANSSEN and A. J. MICHELS. 621.317.755:621.3.029.5/.6

From the fundamental principles of the stroboscopic oscilloscope dealt with in a previous article, to the actual building of an instrument is a big step. The generation of extremely short voltage pulses, the synchronization of those pulses, the construction of the connecting cables and of the variable attenuators for the input voltages are only a few of the problems encountered in the practical execution of this oscilloscope. This second article shows how these problems have been solved.

In a previous article [1]) it was shown how, in principle, an A.F. image of an H.F. voltage v_o can be obtained by causing the latter to be scanned by electrical pulses modulated in phase. We had arrived at a block diagram which is reproduced here (*fig. 1*) but which, as will presently be seen, needs some additional details.

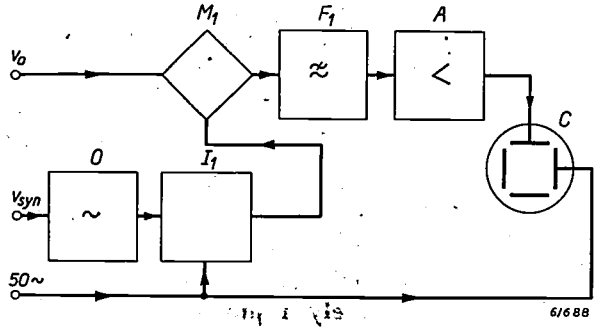


Fig. 1. Simplified block diagram of a stroboscopic oscilloscope. v_o H.F. voltage to be examined, M_1 mixing circuit, F_1 low-pass filter, A A.F. amplifier, C cathode ray tube, I_1 pulse generator, O synchronizing oscillator, v_{syn} synchronizing voltage (synchronous with v_o). The mains voltage (50 c/s) provides both for the phase modulation of the pulses and for the horizontal deflection of the electron beam.

Recapitulating some of the results arrived at in article I, the functioning of the oscilloscope may be described as follows:

The voltage v_o is mixed, in a mixing valve M_1, with pulses derived from a generator I_1. The H.F.-components are removed from the output voltage of the mixer by a filter F_1, and the signal is then amplified by an A.F. amplifier A and applied to the horizontal deflection plates of a cathode ray tube C. The frequency of the pulse generator is controlled by an oscillator O, which in turn is synchronized by an externally applied synchronization voltage v_{syn} which has to be synchronous with v_o. In this way a multiple of the repetition frequency f_i of the pulse is made equal to the fundamental frequency f_o of v_o; as found by calculation in article I, a favourable value for f_i is about 100,000 c/s. The pulses are phase-modulated (preferably sinusoidally) with a frequency ν, which should be chosen as low as possible so as to be able to include in the image as many harmonics of v_o as possible; the mains frequency is used here ($\nu = 50$ c/s). By arranging for the horizontal deflection to take place synchronously with the phase modulation (thus likewise with the mains frequency) a stationary picture is obtained with a linear time scale.

After a brief description of the mixing stage, the filter and the A.F. amplifier, we shall have to describe the pulse generator and its synchronization more fully. Next comes the electronic switch, by means of which two oscillograms can be produced simultaneously, and finally the connecting cables and the construction of the variable attenuators will be dealt with.

It is emphasized that these descriptions refer to an experimental model of the oscillograph, the primary object of which was to put the fundamental theory of the stroboscopic method to a practical test. In tackling the problems encountered no attempt has been made to find the best solution, the only choice being the design which involved the least development work. When dealing with some of the points we shall have occasion to recall this to mind.

The mixing stage

The main elements of the mixing stage are represented in *fig. 2*. The mixing valve is a pentode EF 50.

[1]) J. M. L. Janssen, An experimental "stroboscopic" oscilloscope for frequencies up to about 50 Mc/s, I. Fundamentals, Philips Techn. Rev. 12, 52-59, 1950 (No. 2), here further referred to as article I.

The voltage v_0 to be examined is applied to the control grid and the pulses are fed to the anode. As can be seen, the anode does not receive a positive high tension but a small negative direct voltage (of a few volts) derived from the cathode resistance. The object of this is fully to suppress the anode current in between the pulses, so that the mixing stage is made insensitive for hum voltages and any other interference. The screen grid does receive the usual positive voltage (180 V), which is kept constant by two stabilizing tubes connected in series. It appears to be an advantage to give also the suppressor grid a positive voltage (about 100 volts), for then the conversion conductance is much greater and smaller anode-voltage pulses are needed (15 V amplitude) than when the voltage of the suppressor grid was zero. That is why this grid has been connected to the common point of the stabilizing tubes.

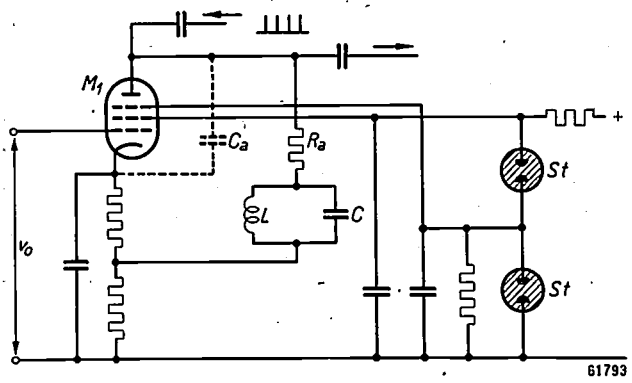


Fig. 2. Mixing circuit. v_0 voltage to be examined, M_1 pentode EF 50, R_a anode resistor, which, together with the circuit L-C and the stray anode capacitance C_a forms a network with flat response curve. St voltage-stabilizing tubes.

To get a high output from the mixing stage a large anode resistance R_a has to be used, but then there is the risk of the stray anode capacitance C_a causing a considerable reduction of the conversion gain at frequencies still below the cut-off frequency of the low-pass filter, which follows the mixing stage. (As calculated in I, this cut-off frequency has to be lower than $^1/_2 f_i \approx 50{,}000$ c/s; in this experimental apparatus we have chosen 40,000 c/s.) Such a falling frequency characteristic would lead to an erroneous reproduction of the v_0 curve. To keep this characteristic horizontal up to 40,000 c/s the well-known method was chosen of connecting in series with the anode resistor R_a an inductor L and a capacitor C in parallel (fig. 2).

The filter and the A.F. amplifier

For the filter we have chosen a simple ladder network with a capacitance in each of the parallel branches and an inductor shunted by a capacitor in each of the series branches. Such a filter has the property that for frequencies above about 50% of the cut-off frequency the transmission is not entirely faithful in phase, with the result that high harmonics of the v_o curve are not exactly reproduced in the right phase relations in the oscillogram. Filter circuits exist (with mutual inductances between the filter stages) which are faithful in phase up to about 85% of the cut-off frequency, but such a complicated filter has not been used in the experimental apparatus.

Following the filter is a push-pull stage (double pentode EFF 51) which with an asymmetrical input (earthed on one side) produces a symmetrical output voltage used for the vertical deflection in the cathode ray tube.

The pulse generator

Differentiating network

There are various ways of generating pulses. For our purpose we have chosen the method whereby the pulses required for the scanning are obtained by using the output voltage of a "differentiating" network fed with a current that suddenly changes at fixed intervals.

In its simplest form such a network consists of two resistors, R_1 and R_2, and a capacitor C_m (*fig. 3a*). When the supply current changes discontinuously by an amount I_0 there is a similar discontinuous variation of the output voltage with the magnitude $I_0 R_1 R_2/(R_1 + R_2)$, since C_m forms a short-circuit for an infinitely rapid change, so that R_1 and R_2 are in parallel at the moment of the sudden change. When, after the sudden change, the input current remains constant then the output voltage v changes according to the expression:

$$v = I_0 \frac{R_1 R_2}{R_1 + R_2} \cdot \exp\left\{-\frac{t}{(R_1 + R_2)\, C_m}\right\}.$$

Thus the shape of this voltage pulse, and also its frequency spectrum, is determined by the product $(R_1 + R_2)C_m$. If a particular frequency spectrum is specified and a given value is chosen for C_m, then the value of $R_1 + R_2$ is fixed. The amplitude of v is then proportional to $R_1 R_2$ and thus is a maximum when $R_1 = R_2$. This maximum becomes greater as C_m is reduced, while keeping $(R_1 + R_2)C_m$ constant.

It would therefore be expected that by a suitable choice of C_m or $(R_1 + R_2)C_m$, pulses could be generated of any amplitude or width. However, no allowance has yet been made for the inevitable stray

capacitances (C_1, C_2, fig. 3*b*) parallel to R_1 and R_2. These capacitances prevent the voltages at R_1 and R_2 from changing discontinuously, the effect being apparent when $(R_1 + R_2)C_m$ becomes of the same order as the time constants R_1C_1 and R_2C_2. This has the effect of widening the pulse and reducing its amplitude.

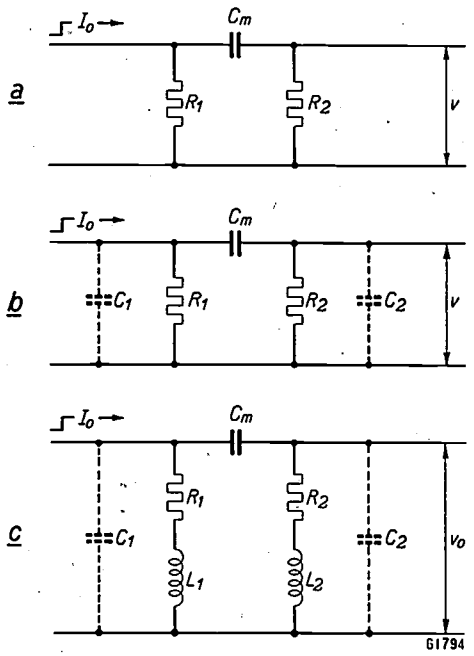


Fig. 3. Differentiating networks. *a*) When the input current is caused to make a step I_o a voltage pulse v arises across the resistor R_2. The coupling capacitance is C_m. *b*) As in (*a*) but with the stray capacitances C_1 and C_2. *c*) To improve the waveform of the pulse the inductors L_1, L_2 are connected in series with R_1 and R_2 respectively.

This effect is of such a nature that for generating pulses as are needed with a stroboscopic oscilloscope for frequencies up to 50 Mc/s — pulses with an amplitude of at least 15 V and a width in the order of at most 10^{-8} s — such simple networks as those of figs. 3*a* and 3*b* are unsuitable. It is possible to counteract the effect of the stray capacitances in the same way as in resistance-coupled amplifiers, by connecting inductors in series with the resistors (fig. 3*c*). Two resonance circuits are then obtained, coupled by the capacitor C_m. Such a system has two characteristic frequencies, both of which are excited when a current surge is sent through one of the circuits. The voltage across the other circuit is then the superposition of two damped oscillations v' and v'', each with one of the characteristic frequencies. It is possible, by a suitable choice of the network elements, to get a situation as follows (*fig. 4*): at the time of the first positive peak of the oscillation with the highest frequency (v' in fig. 4) the other oscillation (v'') is likewise positive, while at the time of the second positive peak of v' the oscillation v'' is negative; owing to the damping the amplitudes of the next peaks are negligible. The negative peaks are of no consequence here. Thus the output voltage $v = v' + v''$ (fully-drawn line in fig. 4) consists mainly of one positive pulse (roughly in the shape of half a sine wave).

Denoting the "average" width of the pulse [2]) by τ ($\approx {}^2/_3$ of the base width) then, as a calculation shows, for the amplitude V_{max} we have approximately:

$$V_{max} \approx \frac{1}{\pi} \cdot \frac{\tau I_0}{C_1 + C_2}.$$

Substituting for V_{max} and τ the values mentioned, viz. 15 V and 10^{-8} s, and for C_1 and C_2 the values that occur in practice ($C_1 = 15$ pF, $C_2 = 7.5$ pF), it is found that the sudden change of the current must be about 100 mA. In practice a higher value proves to be necessary (about 200 mA), partly for the following reason.

So far it has been assumed that the change of the input current is absolutely discontinuous. In point of fact, however, it takes some finite time, with the result that the voltage pulse is wider and also lower than would be the case in the event of an absolute discontinuity. This effect becomes of particular importance as soon as the time taken for the current change becomes of the same order as the permissible duration of the pulse τ.

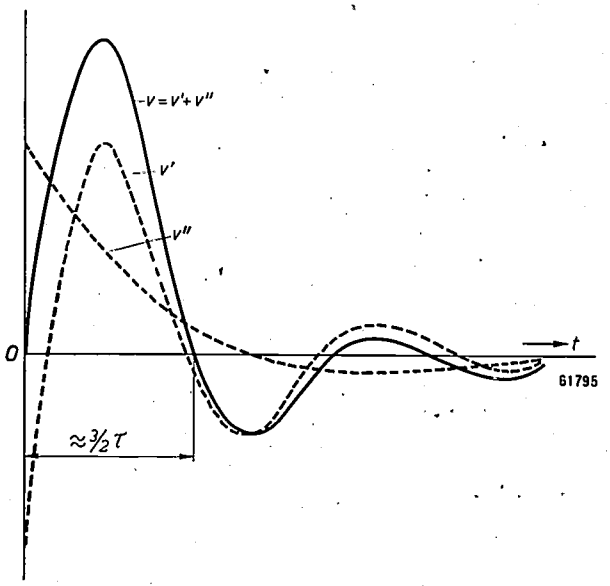


Fig. 4. The output voltage v, as function of the time t, of the network according to fig. 3*c* when a sudden current change occurs at the input. The voltage v is the sum of two characteristic oscillations, v' and v''.

[2]) See I, page 58.

Circuit for producing the sudden current change

This leads up to the question how the change in current can be brought about quickly enough. It is obvious that the differentiating network should be incorporated in the anode circuit of an amplifying valve (P_1, *fig. 5*) and a voltage applied to the control grid of that valve with a curve of such a form that the anode current is suddenly interrupted periodically. Owing to the large anode-current pulses required, the amplifying valve should be an output valve; the type EL 6 for instance can easily produce surges of 200 mA, provided the mean value of the anode current is not too high. This condition can be fulfilled by giving the grid voltage curve a form as represented in *fig. 6a*; the corresponding anode current variation and the form of the pulses supplied by the network are sketched in figs. 6*b* and 6*c* respectively. The manner in which this grid voltage is obtained will be explained in the next section.

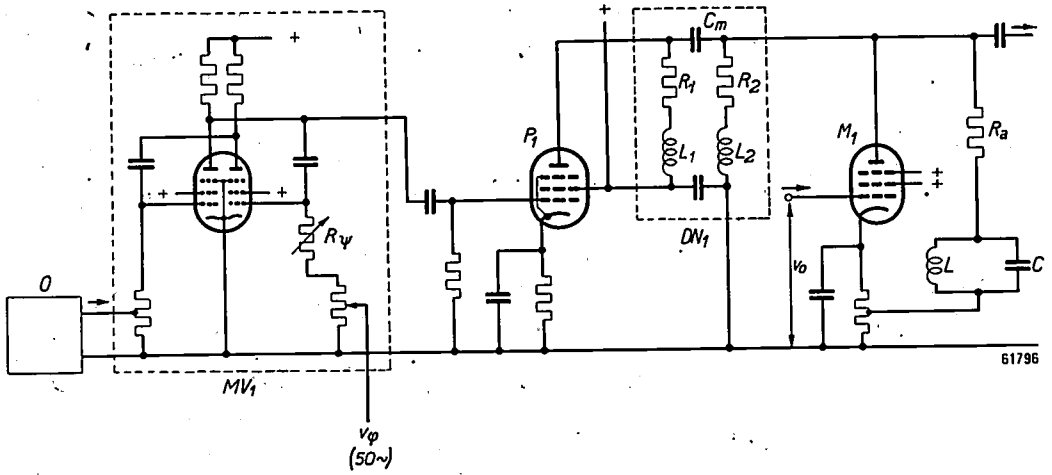


Fig. 5. The pulse generator I_1 (fig. 1) consists of a multivibrator MV_1, an output valve P_1 and a differentiating network DN_1 according to fig. 3c. O is the synchronizing oscillator. M_1, with R_a, L and C, forms the mixing circuit according to fig. 2. The mean position of the pulse can be adjusted with the resistor R_ψ. The voltage v_φ provides for the phase modulation. The valve in the multivibrator is a double pentode EFF 51, P_1 is an EL 6 valve.

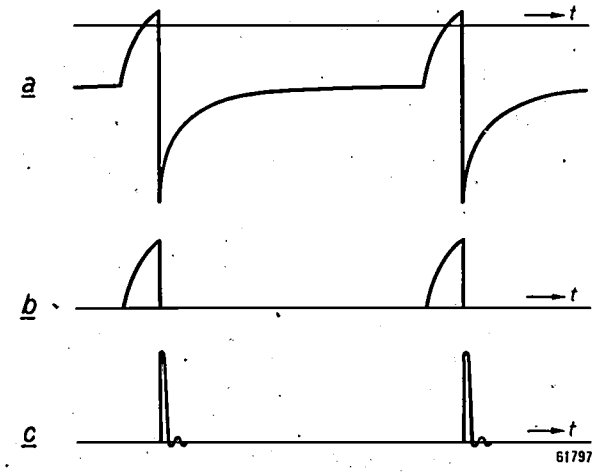


Fig. 6. Plotted as a function of the time t: *a*) the voltage on the control grid of the valve P_1 (fig. 5), (*b*) the anode current of that valve, and (*c*) the pulse-shaped output voltage of the differentiating network DN_1 (fig. 5).

Multivibrator

A voltage as indicated in fig. 6*a* can be obtained with the aid of a multivibrator (a set of two valves with the control grid of one coupled to the anode of the other and vice versa, the adjustment being so chosen that current is passed by the valves alternately).

In fig. 5, MV_1 represents the system used in this experimental apparatus. This multivibrator has one valve of the EFF 51 type, containing two pentode systems. A sinusoidal voltage derived from the oscillator O (to be described later) is applied to one control grid of this valve (cf. fig. 1); this voltage synchronizes the multivibrator, a multiple of the repetition frequency f_i (about 100,000 c/s) of the pulses obtained being made equal to the fundamental frequency of the voltage v_o that is to be examined. To the other control grid of the valve EFF 51 a voltage v_φ of 50 c/s is applied, which produces the phase modulation of the pulses that is necessary for scanning; this voltage, inter alia, determines the moment at which the current switches over from one pentode system to the other. The amplitude of the 50 c/s voltage is variable, so that the phase sweep of the pulse, i.e. the size of the part of the v_o curve displayed on the screen, can be adjusted.

A portion (R_φ, fig. 5) of the resistance connected in series with the grid to which the 50 c/s voltage is

fed is made variable for adjusting the mean position of the pulses with respect to v_o, and thus the point of the v_o curve which forms the centre of the oscilloscope waveform.

The working of the pulse generator can be summarized as follows. A multivibrator supplies a voltage showing a step with a frequency of about 100,000 c/s. On being applied to the grid of an EL 6 valve this voltage causes a periodical change of the anode current from 200 mA to zero, thus giving rise to periodical voltage pulses at the output of a differentiating network, the average width of these impulses being in the order of 10^{-8} s. In the mixing circuit the pulses scan the voltage v_o which is being examined.

Synchronization

We now come to the description of the system supplying the voltage for the synchronization of the multivibrator. This voltage has to have a frequency round about 100 kc/s and the fundamental frequency f_o of the voltage v_o to be examined has to be an exact multiple of this frequency.

The voltage in question is derived from an oscillator O (*fig. 7*), the frequency of which is controlled by a reactance valve RT, which is shunted across the oscillating circuit of the oscillator via its anode and cathode and is so connected that, with respect to the oscillating circuit, it behaves as an inductance or a capacitance, the value of which can be continuously controlled by varying its control grid bias [3]).

In order to control the oscillator frequency automatically a device is needed which supplies a correcting voltage to the reactance valve as soon as the oscillator frequency deviates from the right value. This device is built up from elements similar to those we have come across before, namely a mixing circuit M_2 (fig. 7), in which, after having undergone a variable attentuation and a constant amplification, the synchronizing voltage v_{syn} is mixed with pulses generated by a pulse generator I_2, which is likewise controlled by the oscillator O. This pulse generator also consists of a multivibrator (MV_2) followed by an output valve (P_2) with a differentiating network (DN_2) in the anode circuit. This network supplies voltage pulses with a repetition frequency equal to the oscillator frequency. The main difference compared with the pulse generator I_1 previously described is that in I_2 the pulses are not modulated in phase.

The D.C. component of the voltage across the anode resistor of the mixing valve M_2 serves as control voltage for the reactance valve.

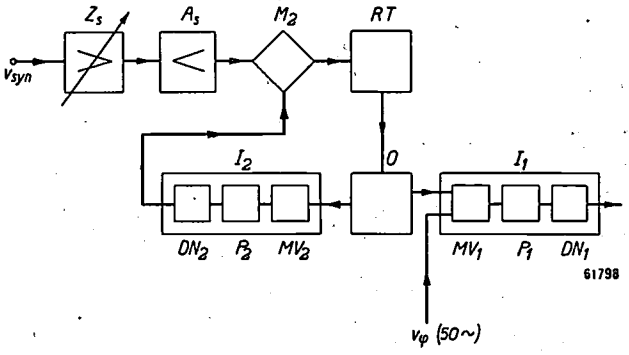


Fig. 7. Synchronization circuit. The oscillator O controls not only the pulse generator I_1 (fig. 1) but also a pulse generator I_2, the pulses from which are mixed with the synchronizing voltage in a mixing circuit M_2. The result is a direct voltage which changes in the event of failure of the synchronization. This direct voltage controls a reactance valve RT which corrects the oscillator frequency when necessary. MV_2, P_2, DN_2: multivibrator, output valve and differentiating network forming the pulse generator I_2, similar to I_1. The externally applied synchronization voltage v_{syn} reaches the mixing valve via a variable attenuator Z_s and an amplifier A_s.

Let us suppose that a multiple of the pulse frequency of I_2 is indeed exactly equal to the frequency f_o of the voltage on the control grid of the mixing valve M_2. The anode current of this valve then consists of pulses equal in amplitude (see fig. 2*b* of article I). As soon as the oscillator frequency begins to change, the phase of the pulses with respect to the said alternating grid voltage is no longer constant; the anode-current pulses of M_2 and thus the control voltage of the reactance valve change in value, so that this valve thereby corrects the oscillator frequency.

The system described here is a special case of what is known as the "I.G.O." system (= *I*mpulse *G*overned *O*scillator) used in transmitters and receivers for synchronizing oscillators. In that system the synchronizing voltage usually has a frequency (from a quartz crystal) higher than the oscillator frequency. In our case it is just the other way round (at most the two frequencies are equal).

The range of synchronization is greater if the amplification in the circuit formed by the oscillator O, the pulse generator I_2, the mixing circuit M_2 and the reactance valve RT is large. When the amplification exceeds a certain threshold value instability arises, as is the case in any regulating system. In order to make the amplification as high as possible without reaching the point of instability it

[3]) The principle of a reactance valve has been described several times in this journal; see, e.g. Philips Techn. Rev. 8, p. 47 (fig. 7) or p. 122, 1946.

[4]) H. B. R. Boosman and E. H. Hugenholtz, Frequency control in transmitters, Communication News 9, 21-32, 1947.

is necessary to provide, inter alia, for the best possible phase equality, i.e. a tight coupling, between the oscillator O and the pulse generator I_2. This coupling is brought about by means of a diode, so that only the peaks of the sinusoidal oscillator voltage have a synchronizing action upon the pulse generator [5]).

Synchronism ($nf_i = f_o$) occurs when, instead of a continuous strip of light in which nothing can be distinguished, a stationary picture is formed on the screen of the cathode ray tube. When the capacitance in the oscillating circuit of the oscillator is changed it depends both upon the fundamental frequency f_o and upon the amplitude of the synchronization voltage whether isolated or connected synchronization ranges are found for the successive sub-multiples of f_o. When $f_o > 1$ Mc/s one can always find continuous synchronization ranges (and thus the capacitor need not be turned) merely by giving the synchronization voltage a suitable value, which can be done by means of a variable attenuator to be described later (Z_s, fig. 7). If $f_o < 1$ Mc/s then only isolated synchronization ranges occur, which have to be found by turning the capacitor.

Due to the amplifier A_s (fig. 7) a synchronizing input voltage v_{syn} of low value (about 20 mV) suffices. This H.F. amplifier contains one EF 50 valve with a resistor in the anode circuit; at frequencies higher than about 1 Mc/s the amplification rapidly diminishes as the frequency rises, owing to the stray anode capacitance. This is permissible in this case because the higher the frequency f_o of v_{syn}, the closer the sub-multiples of f_o round about 100,000 c/s lie together, so that adjacent synchronization ranges can be obtained with a smaller synchronizing voltage.

The electronic switch

As already noted in article I, the phase modulation of the pulses can be used for producing two oscillograms simultaneously by causing the pulses to scan on the return stroke a voltage curve which differs from that scanned on the forward stroke. This greatly facilitates, for instance, measuring of the phase difference between two voltages.

All that is needed for this purpose is an electronic switch, which can be of very simple construction in this case (*fig. 8*). The input voltages v_{oI} and v_{oI_I}

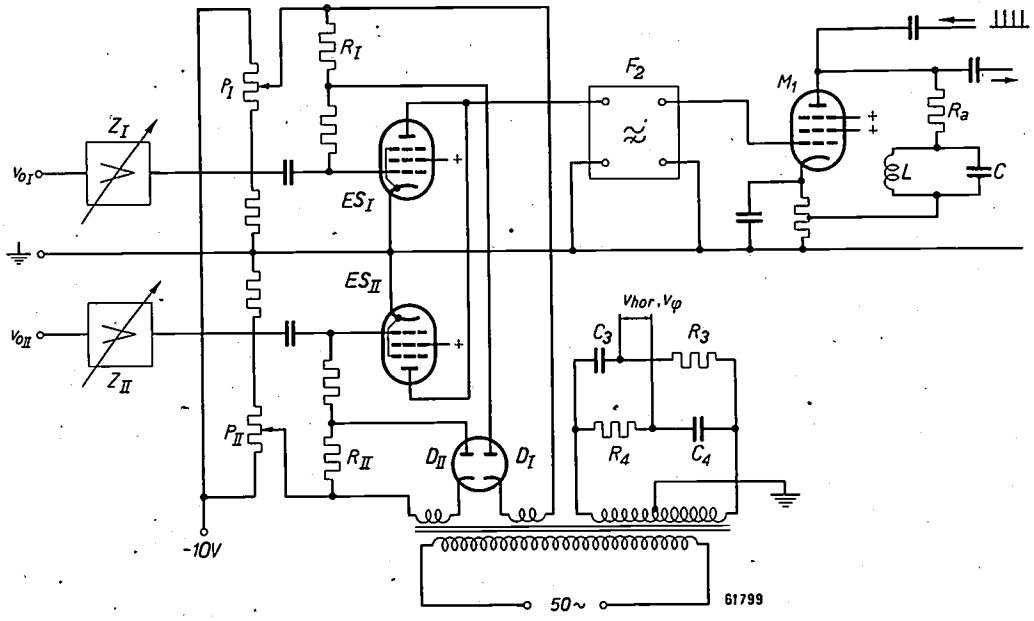


Fig. 8. Electronic switch. The two input voltages to be viewed, v_{oI} and v_{oII}, are fed, each via an attenuator Z_I, Z_{II}, to the input of the valves ES_I and ES_{II} respectively. These valves are alternately blocked by the semi-sinusoidal switching voltages occurring across the resistors R_I, R_{II} via the diodes D_I, D_{II}. The voltage $v_{hor} = v_\varphi$, which brings about both the horizontal deflection and the phase modulation, is shifted 90° in phase in the network C_3-R_3-C_4-R_4 with respect to the alternating voltages, which in turn block the valves ES_I and ES_{II}. P_I, P_{II} potentiometers for adjusting the bias, F_2 high-pass filter; M_1 with R_a, L and C: mixing circuit according to fig. 2. The valves ES_I and ES_{II} are both of the EF 50 type, the double diode D_I-D_{II} is an EB 4 valve.

[5]) For a more detailed treatise on synchronization reference is made to J. M. L. Janssen, A cathode-ray oscillograph for periodic phenomena of high frequencies, Philips Res. Rep., 5, 205-240, 1950 (No. 3).

are fed, via a variable attenuator, to the control grids of the valves ES_I and ES_{II} respectively (both of the type EF 50) in the circuit of the electronic switch. Switching over is done with the mains frequency, at the extreme values of the phase sweep of the pulses. For this purpose the alternating voltage v_φ bringing about the phase modulation is in quadrature with the alternating voltages fed to the electronic switch; the phase shift of 90° is obtained with the aid of capacitors and resistors (C_3, R_3, C_4, R_4).

The electronic switch works as follows. During the half cycle of the mains voltage when the diode D_{II} is conducting the valve ES_{II} is blocked by the voltage across the resistor R_{II}; the diode D_I and the resistor R_I are then non-conducting, the valve ES_I functions normally and the voltage v_{oI} is reproduced on the screen of the cathode ray tube. During the next half cycle the reverse takes place: current flows through D_I and R_I, ES_I is blocked, ES_{II} functions normally and v_{oII} is displayed.

The fact that no rectangular voltages are needed for the blocking of the valves and that semi-sinusoidal voltages suffice is due to the fact that a fairly long time is available for switching over at each peak of the phase sweep, so that the valves need not be cut off suddenly.

Between the common output of the valves ES_I and ES_{II} and the input of the mixing circuit M_1 is a high-pass filter F_2 cutting out any low-frequency voltages arising from the switching.

In addition to the attenuators Z_I and Z_{II}, which are variable in steps and are described below, continuous control is provided in the form of the potentiometers P_I and P_{II}, with which the negative control grid bias of the valves ES_I and ES_{II} can be adjusted.

The connecting cables and the attenuators

When measuring or examining with an oscilloscope voltages with very high frequency components it cannot be left to chance what kind of leads are used for connection to the voltmeter or oscilloscope: in the first place, owing to the capacitance and the damping which lie in parallel with the voltage source when connecting up, the voltage might change considerably, and moreover stationary waves might arise in the lead itself, so that the voltage at one end of the lead would differ considerably from that at the other.

These factors have already been discussed in connection with a millivoltmeter for frequencies up to 30 Mc/s described in this journal [6]). In the case of the oscilloscope described the problem is still more difficult because the transmission along the cable has to be faithful not only in amplitude but also in phase and, moreover, it is desired to extend the frequency limit well beyond 30 Mc/s.

As to the transmission the best solution would be to terminate the cable at the end with a resistance equal to its characteristic impedance. The input impedance of the cable would then likewise be of this value, which, however, would not exceed about 100 ohms and thus would form a very heavy load on the voltage source. In order to avoid this, the cable could be fitted with a "probe" containing an amplifying valve. This valve could be advantageously connected as a cathode follower to match the low impedance of the cable. Furthermore, the probe would have to contain a variable attenuator preceding the valve if voltages are to be examined which, unattenuated, would overload the valve.

Since, among other reasons, our object in building this oscilloscope was in the first place only to verify the accuracy of the principles, we have confined ourselves to fitting an attenuator in the probe, namely a capacitative attenuator. In principle this attenuator — like the one discussed in the article quoted in footnote [5]) — could consist of a variable series capacitor, but in this case the transmission characteristic of the cable would then vary considerably with the value of this capacitor, and thus with the attenuation. For a constant transmission characteristic it is necessary that the input capacitance, viewed from the cable, is constant. This has been achieved by using as attenuator a number of capacitors connected as indicated in *fig. 9*. The voltage to be examined is connected between *0* and *1*, with

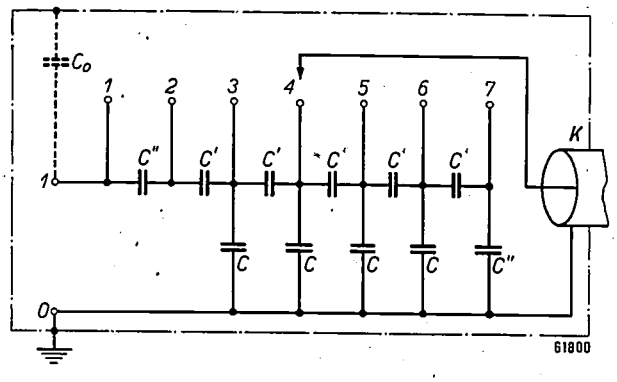
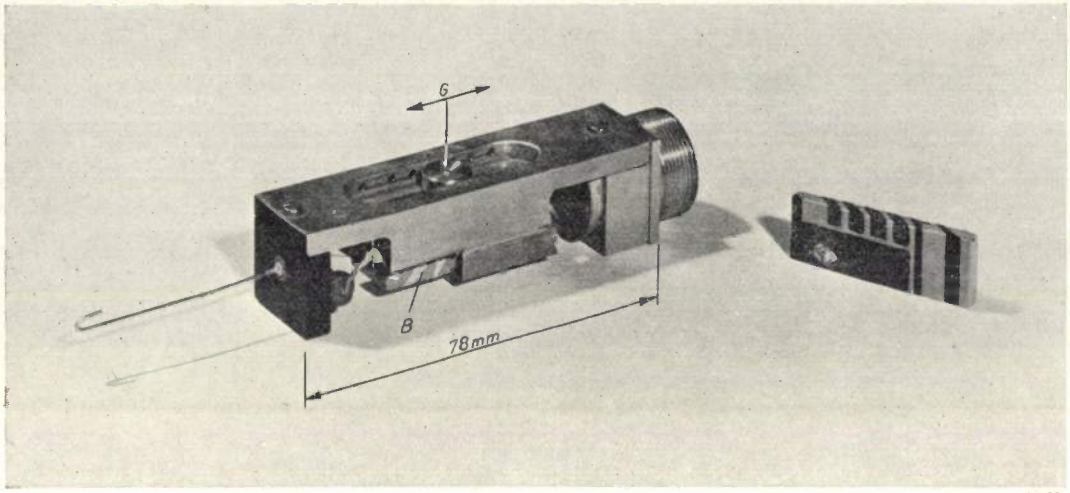


Fig. 9. Network of capacitors acting as variable attenuator, with the property that, with respect to the cable K, it behaves as a capacitance that is independent of the tappings *2* ... *7* to which the cable core is connected (this does not apply for tapping *1*, where the cable is connected directly to the input terminals *0-1*). The capacitances used, whereby the attenuation can be adjusted in steps of $\sqrt{10}$, are: $C = 10$ pF, $C' = 6.6$ pF, $C'' = 14.6$ pF. C_0 is the stray capacitance with respect to earth.

[6]) H. J. Lindenhovius, G. Arbelet and J. C. van der Breggen, Philips Techn. Rev. 11, 206-214, 1949 (No. 7).

the cable core connected to one of the points *1* ... *7*; in this order the attenuation increases step by step by a factor $\sqrt{10}$. Except in the position *1*, where the cable is connected direct to the measuring point and thus the voltage to be examined is transmitted dielectric constant and with the eight capacitor plates made by local plating with silver. A sliding contact runs over silver contact points soldered onto the plates (except the earthed one). The whole is mounted in a probe of small dimensions (*fig. 10*).



61792

Fig. 10. On the right a block of ceramic material with silver plates together forming the network of capacitors drawn in fig. 9. On the left the complete attenuator (opened); *B* is the ceramic block, *G* the knob of the sliding contact. The right-hand end is connected to the cable; the hooked wire on the left is contacted with the point carrying the high-frequency voltage that is to be examined.

unattenuated, the input capacitance, viewed from the cable, is constant (19.3 pF in the case of fig. 9) and for every step of attenuation the system has the same transmission characteristic.

At 45 Mc/s this characteristic shows a peak, but it has been possible to limit this to 1.25 times the height of the flat part by connecting a damping resistor to the end of the cable. This remaining boosting is favourable in counteracting the decline in sensitivity arising in this frequency range as a consequence of the finite pulse width (see article I).

In order to minimize the input capacitance of the oscilloscope (with cable and attenuator), a cable with a very low capacitance (13 pF per metre cable length) has been used, this being obtained by making it with a core (of tungsten for strength) only 50 μ thick. In the positions *4* ... *7* the input capacitance (between *1* and *0*) is about 5 pF, in the positions *2* and *3* it is somewhat higher, and in position *1* — due to the cable capacitance and the capacitance of the first valve being connected in parallel — the input capacitance is 30 to 35 pF.

The eleven capacitors of the attenuator have been assembled in a convenient shape in the form of a rectangular block of ceramic material with high

Results

A picture of the experimental model of the stroboscopic oscilloscope is given in *fig. 11*, with an explanation of the controls in the subscript.

In *fig. 12* the sensitivity *s* (deflection on the screen in relation to the input voltage) of the instrument including the cable has been plotted on a relative scale (maximum taken as 1) and also the phase shift α in the cable, both as functions of the frequency. The decline of *s* at frequencies higher than 30 Mc/s — which without the boosting effect in the cable would be still greater — is due to the finite pulse width. Up to that frequency level the phase characteristic of the cable is practically rectilinear. This limit could be raised to higher frequencies by using a cable terminated with the characteristic impedance, as discussed above.

As a consequence this oscilloscope is of universal use for examining voltages containing harmonics up to 30 Mc/s. The possibility of "microscopic scanning" is recalled to mind, whereby any part of a cycle can be viewed over the full width of the screen, so that a particular detail can be studied. When one has to do with sinusoidal voltages — so that no attention need be paid to the phase —

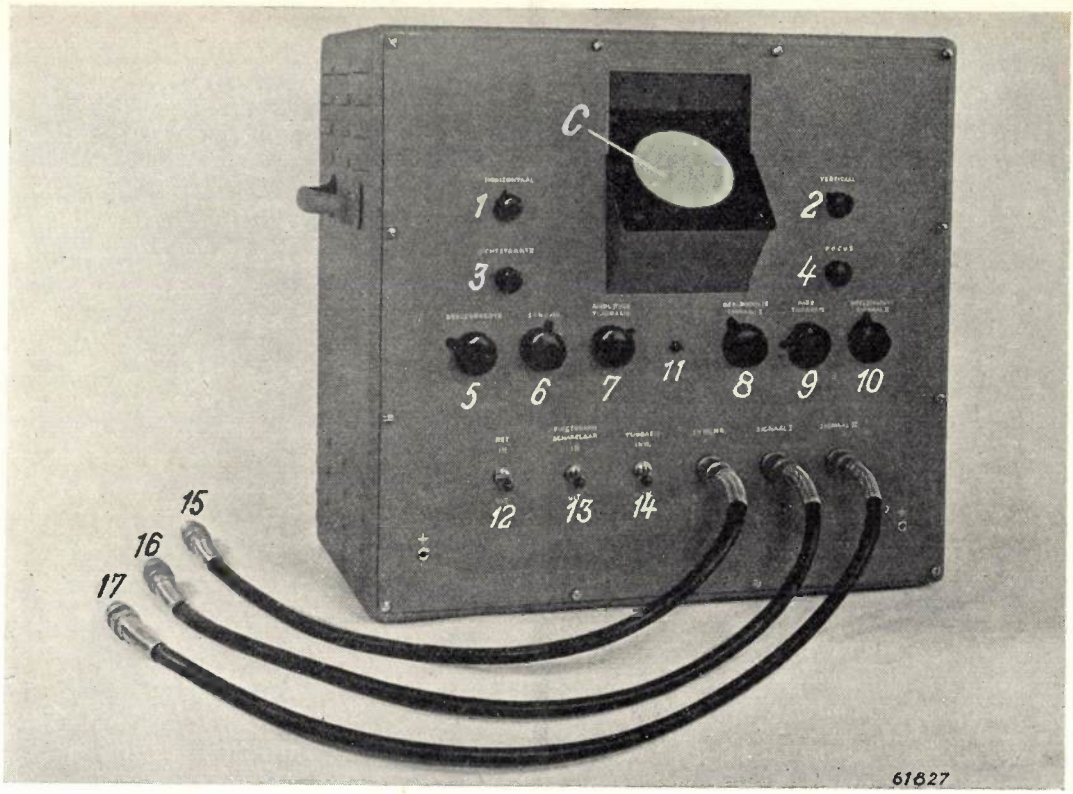


Fig. 11. Experimental model of the stroboscopic oscilloscope. On either side of the cathode ray tube *C* (screen diameter 9 cm) are the knobs *1* and *2* for the horizontal and vertical displacement of the oscillogram, knob *3* for adjusting the brightness (current intensity of the beam) and knob *4* for the focusing. Underneath are the following six controls: *5* adjustment of picture width (amplitude of v_{hor}), *6* synchronization (capacitor of oscillator *O*, fig. 7), *7* size of the part of signal scanned ($\Delta\varphi$, adjustable with the amplitude of v_φ, see fig. 5), *9* selection of the centre of the scanned part (resistor R_ψ, fig. 5), *8* and *10* continuous attenuators for the input voltages (potentiometers P_I and P_{II}, fig. 8). *11* is a pilot lamp. At the bottom on the left are three switches: *12* for the mains, *13* for the electronic switch, whilst with *14* any external voltage can be applied as time base. *15*, *16* and *17* are the cables (here without attenuators) for the synchronizing voltage and the two voltages to be examined.

this instrument is still quite useful up to much higher frequencies, namely to about 70 Mc/s, although sensitivity will then be somewhat lower. It can render excellent service in measuring, up to this frequency limit, gain and phase diagrams of amplifiers (for the latter it is only necessary that the two inputs have the same phase characteristic).

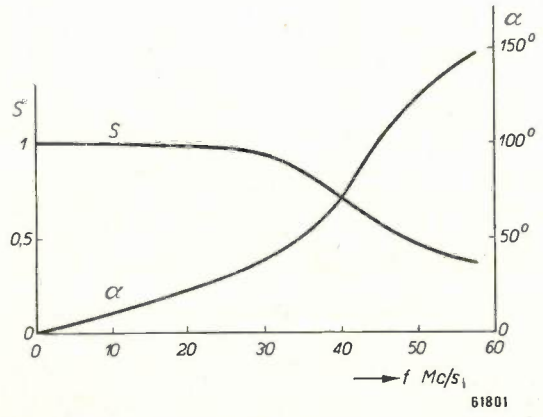


Fig. 12. Relative sensitivity *s* of the complete instrument (including cable) and the phase shift α in the cable, as functions of the frequency *f*.

The lower limit of the frequency range for which the oscilloscope can be used is in any case not higher than 1 Mc/s, above which limit, with a synchronization voltage of 20 mV or more, adjacent synchronization ranges are found and the phase sweep of the pulses can be made large enough for scanning more than one cycle if so desired. Between 1 and 0.1 Mc/s however, as already stated, there are isolated synchronization ranges in which the oscilloscope can operate. In that case less than one cycle is scanned; for a larger phase sweep the phase modulation of the pulses would have to be brought about in a more complicated manner than has been judged sufficient for the experimental model, the behaviour of which at high frequencies was of most interest to us.

Finally it is to be mentioned that the maximum absolute sensitivity is about 1 cm picture height per 20 mV (R.M.S. value) voltage.

Summary. A description is given of the principal elements of an experimental model of the stroboscopic oscilloscope, the fundamental principles of which were explained in a previous article. These elements are: the mixing circuit, the low-pass filter and the A.F. amplifier, the pulse generator, the oscillator which, together with the auxiliary circuits, keeps the pulse generator synchronized, the electronic switch which allows of the simultaneous displaying of two signals, the connecting cables and the attenuators. With the oscilloscope described here non-sinusoidal voltages with harmonics up to 30 Mc/s can be reproduced practically undistorted. Details of a non-sinusoidal voltage can be examined over the full width of the screen. Phase and amplitude characteristics of amplifiers, for instance, can be measured up to frequencies of about 70 Mc/s. The sensitivity is about 1 cm picture height per 20 mV (R.M.S.) input voltage.

CONTROLLING THE LUMINOUS INTENSITY OF FLUORESCENT LAMPS WITH THE AID OF RELAY VALVES

by K. W. HESS and F. H. de JONG.

621.327.42 : 621.3.077.64 : 621.385.38

Gas-discharge lamps, such as tubular fluorescent lamps, could not hitherto be successfully applied in places where the luminous intensity has to be continuously adjustable within wide limits, as for instance in theatres. Relay valves (thyratrons) however offer a good solution of this problem, a solution which shows the way to new possibilities of application for these lamps.

Controlling the luminous intensity of incandescent lamps

In some cases it is necessary that the luminous intensity of lamps should be continuously variable from the maximum down to a small fraction and vice versa. An example of such a case is the gradual lowering and raising of the lighting in theatres, cinemas and lecture halls and also stage lighting. Further, there are certain systems of show-window lighting and illuminated advertising signs where it is desired to vary the luminous intensity in order to attract attention.

When incandescent lamps are used there are a number of solutions which in essence are very simple. The oldest is the use of resistors, either as series resistors or as potentiometers, for varying the current flowing through the lamps. This method is still largely applied and for small installations provides the solution involving the least initial cost, but for large installations such resistors may be rather expensive, partly on account of the high demands that have to be met in theatres and the like from the point of view of safety against fire. Moreover, the cost of the power lost in the resistors is an item of consideration, at least if the installation has to serve not only for gradually lowering and raising the lighting at the beginning and end of the performances but also for dimming the lighting for long periods.

More economical and much less dangerous, from the point of view of fire risk, are variable ratio transformers [1]) the secondary voltage of which can be varied by means of tappings or a sliding contact running over the turns of the secondary coil. In modern installations this is the solution mostly chosen. The advantages mentioned apply also to induction regulators or rotary transformers and to transductors (D.C. controlled A.C. chokes). The latter possess the additional advantage that they lend themselves well for remote control without the intermediary of mechanical transmissions or servo-motors, an advantage that is particularly of importance for hall and stage lighting installations.

[1]) Cases where only a D.C. supply is available are left out of consideration here.

Difficulties encountered in controlling TL lamps

For incandescent lamps these solutions have in course of time been developed to a high degree of perfection. Nowadays we have also tubular fluorescent lamps, which for many applications are to be preferred to incandescent lamps for various reasons. The most important of these reasons are the higher efficiency and longer life of these lamps, further their shape and the spectral composition of the light they produce, allowing of certain decorative effects, and finally the ease with which the small amount of heat generated is carried off. Now how do matters stand when one tries to use these lamps in places where the luminous intensity has to be gradually varied? It is to be foreseen that in the case of fluorescent lamps, through which the current does not begin to flow until a certain voltage — the ignition voltage — has been exceeded, any attempt to vary the luminous intensity, for instance by lowering the voltage, must lead to difficulties which do not arise in the control of incandescent lamps. These difficulties we shall deal with presently, but first attention has to be drawn to a matter which, regardless of the method of control to be employed, presents itself in the case of TL lamps and is related to the fact that, like most gas-discharge lamps, these lamps are fitted with incandescent cathodes. These cathodes are so dimensioned that in normal use they are kept at the right temperature by the discharge itself (ionic bombardment). If, in order to reduce the luminous intensity, the current flowing through the lamp is lowered then the temperature

of the cathodes drops. This is detrimental to the life of the lamp and moreover, since the ignition voltage is thereby raised, it is not conducive to steady burning. If the luminous intensity — thus the current — of these lamps is to be controlled it is above all necessary to supply heating current to both cathodes, for which purpose a separate filament transformer with two separate secondary coils is needed for each lamp. We shall revert to this point later.

It is understandable that for varying the luminous intensity of TL lamps the known methods developed for incandescent lamps were tried out first, though to little effect, as will be seen from the following.

Let us first take the case of a series resistor. For the luminous flux to be reduced to a certain fraction of the maximum a much larger resistor is needed for a TL lamp than for an incandescent lamp of the same power, because with an incandescent lamp a relatively very small drop in temperature of the filament is sufficient to reduce the luminous flux to a certain extent and this temperature, in turn, rapidly decreases with the lowering of the current passing through the lamp. Consequently a rather small series resistor, say of twice the resistance of the filament in the hot state, suffices to reduce the luminous flux of an incandescent lamp approximately to zero; for a 40 W, 220 V lamp, for instance, about 2500 ohms. In the case of a TL lamp, however, the luminous flux is approximately proportional to the current flowing through the lamp; thus it diminishes only approximately in inverse proportion to the series resistance. In order to reduce the current of a 40 W TL lamp so far as to lower the luminous flux almost to zero the series resistance has to be raised to something like 100,000 ohms, thus many times greater than the series resistance of the corresponding incandescent lamp.

Even if it were possible to overcome this objection of such a large series resistor a difficulty would still be encountered, at least with the TL lamps in their original design, in connection with the high re-ignition voltage of these lamps. To understand how the re-ignition voltage gives rise to difficulties in this connection it is necessary to consider the working of a TL lamp connected in the normal way in series with a choke to an alternating voltage of 220 V (*fig. 1*).

The manner in which the lamp is ignited by a starter has been described previously in this journal [2]). In the stationary state there is an arc voltage of about 110 V across the lamp. The discharge current i, which lags with respect to the mains voltage, passes through zero without remaining at that level a finite length of time. Thus there is no interval of zero current.

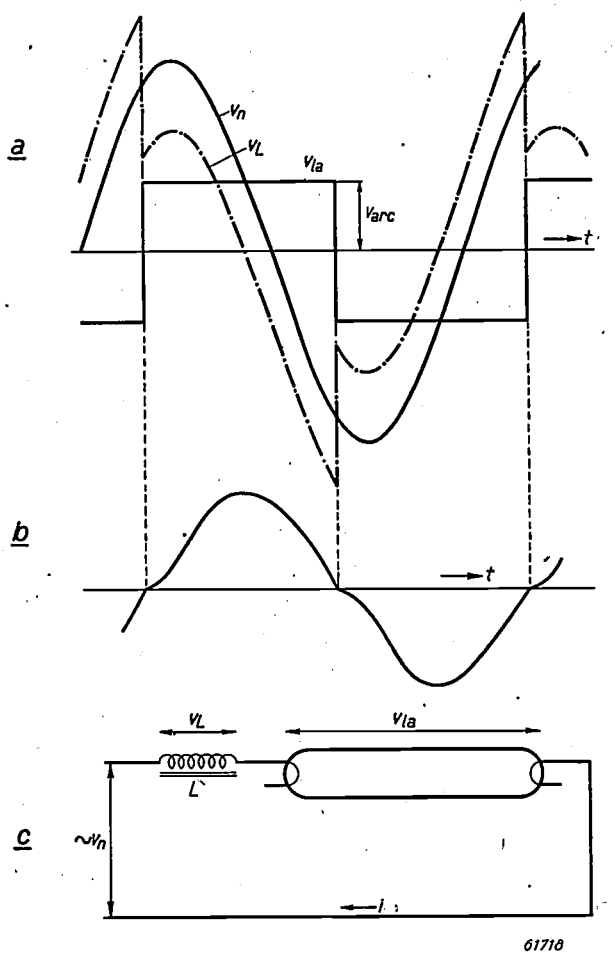


Fig. 1. Voltages and current of a normally working TL lamp (40 W) connected in series with a choke L to a mains voltage of 220 V. Plotted as functions of the time t are: (*a*) the mains voltage v_n, the voltage v_{la} across the lamp, and the voltage $v_L = v_n - v_{la}$ across the choke, and at (*b*) the current i. No zero current interval occurs here. v_{arc} = arc voltage. (*c*) is the circuit diagram. (The curves for v_{la} and v_L are somewhat schematically drawn.)

When the choke is replaced by a resistor the situation is different, for then the current already drops to zero (at $t = t_1$, *fig. 2*) when the mains voltage has fallen to the arc voltage, thus before the mains voltage becomes zero. In the reverse direction the current cannot begin to flow until the mains voltage has changed its polarity and, at $t = t_2$, has reached a level — the re-ignition voltage — which, as we shall presently see, is much higher than the arc voltage. The result is the occurrence of a zero current gap twice in every cycle (also sometimes referred to as the dark period, though during this gap the TL lamp continues to yield some light owing to the after-glow of the fluorescent substan-

[2]) Th. Hehenkamp, A rapid-action starter switch for fluorescent lamps, Philips Techn. Rev. **10**, 141-149, 1948.

ces). Now such a gap is undesirable, because the ions present in the lamp at the beginning of a gap rapidly diminish in number owing to recombination, and the lower the ionic concentration the higher is the re-ignition voltage. In the aforementioned case of a normally working lamp connected in series with a choke, where no such gap occurs, upon the current being reduced to zero the ionic concentration is so great that the re-ignition voltage is but little higher than the arc voltage. With a resistor, however, where a zero current gap always occurs, we have to do with a much higher re-ignition voltage. If, by increasing the resistance, we reduce the current then the arc voltage is raised slightly (a property of the gas discharge in the TL lamp), the zero current gap becomes longer and the re-ignition voltage is raised still higher. With the TL lamps of the old design this increase of the re-ignition voltage may in fact be so great that the lamp does not ignite at all, so that when the luminous intensity of the lamp is reduced there is a risk of the lamp extinguishing, suddenly or after some flickering, and not being able to re-ignite without the help of a starter. Gradual adjustment of the light is then possible only between the maximum luminous intensity and a level that is but little below it. As far as this latter level is concerned there may be great differences between various specimens of the same type of lamp, so that the effect in the lighting of an auditorium would be erratic and unsatisfactory.

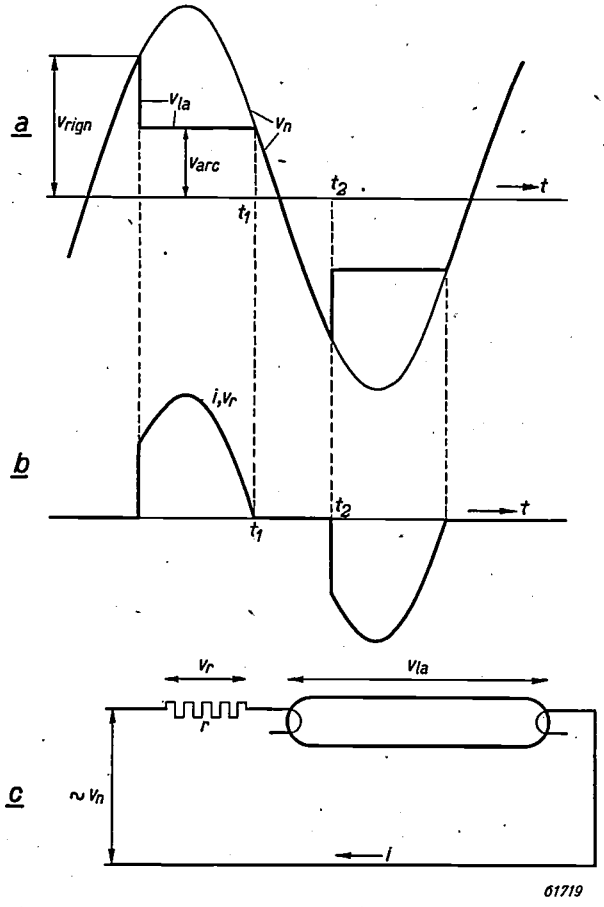


Fig. 2. Voltages and current of a TL lamp connected in series with a resistance r to an alternating voltage of 220 V. v_r = voltage across the resistor; v_{rign} = re-ignition voltage; the meanings of the other symbols are as in fig. 1. Here there is indeed a zero current interval (t_1-t_2).

The situation is not much better when the lamp is provided with the normal choke (as is desired anyhow in order to avoid dissipative losses while the lamp is burning at full strength) and the variable series resistor is used only for varying the current [3].. At low levels of lighting (large series resistance) the choke has but little effect and the troubles just mentioned still occur.

Much more favourable is the behaviour of the modern type of TL lamps on the market several years already. Thanks to a conducting strip on the glass (to which we shall revert later) these lamps ignite more readily than lamps without this strip. With a series resistance of the order of 100,000 ohms the luminous intensity of such a lamp can indeed be gradually controlled between the maximum strength and a very low level. But this does not mean to say that the problem has thereby been solved. So far we have been considering the case of only one TL lamp, but of course in a hall there is a large number of lamps and one cannot deal with them in the same way as with incandescent lamps by employing one common variable resistor for a large number of lamps connected in parallel. The fact of the matter is that as the number of igniting TL lamps increases after switching on (they do not all ignite exactly at the same instant) the voltage drop in the common resistor increases until ultimately the voltage remaining for the rest of the lamps drops below the ignition level. Each lamp would therefore require its own variable resistor and in a large installation this would complicate matters enormously.

The objections set forth here against the use of series resistors also hold in the case of variable chokes (e.g. transductors) being applied.

As to the other solutions that can satisfactorily be applied with incandescent lamps, these all amount to a variation of the amplitude of the alternating voltage applied (variable ratio transformers, induction regulators, resistors connected as potentio-

[3]) See: The dimming of fluorescent lamps, Electrical Times **115**, 641, 1949 (No. 3001), or Fluorescent stage lighting, Light and Lighting **42**, 169-170, 1949 (No. 7).

meters). If this were to be applied to TL lamps then upon the voltage being lowered the re-ignition level would soon no longer be reached and the lamps would suddenly extinguish.

The foregoing accounts for the fact that hitherto little use, if any, has been made of TL lamps for the lighting of halls and other places where gradual dimming of the lights is desired. The same applies for gas-discharge lamps for high tension (with fluorescence or without, such as neon tubes), which in such cases could otherwise be used to good purpose for decorative reasons.

Control with relay valves

Apart from the methods summed up in the foregoing there is an entirely different way of controlling an alternating current, namely by means of relay valves (thyratrons). These are gas-filled rectifying valves fitted with a control grid. Briefly the working of these valves [4]) amounts to this, that with a positive anode voltage the valve ignites only when the grid voltage exceeds a certain critical value (in general this critical value is a function of the anode voltage.) Once the valve has been ignited the magnitude of the grid voltage no longer has any effect upon the anode current. This current drops to zero only when the anode circuit is broken or, as is always the case in a circuit fed with alternating voltage, when at a certain moment the anode becomes negative.

When the grid of a relay valve is kept sufficiently negative all the time, while the anode circuit is fed with an alternating voltage, then no current at all passes. If the grid voltage is kept continuously above the critical value, or if it is arranged to rise above that level at the moment that the anode becomes positive, then the valve allows current to pass during the longest possible intervals. Assuming, for the sake of simplicity, that there is only a resistance in the anode circuit, then these intervals are approximately half-cycles [5]). Within these two extremes a continuous change is possible by displacing the instant of ignition within such a half-cycle. This can be arranged, for instance, by having a grid voltage consisting partly of an alternating voltage that can be changed in phase with respect to the

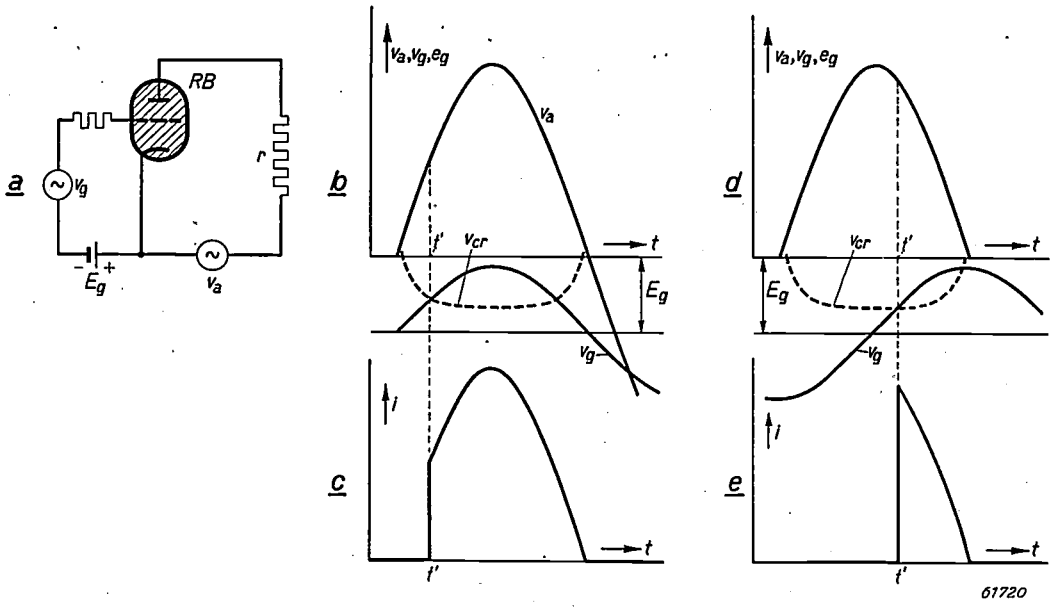


Fig. 3. *a*) Relay valve (thyratron) *RB*, of which the A.C. grid voltage v_g can be shifted in phase with respect to the A.C. anode voltage v_a. The valve is loaded with a resistor *r*. E_g = grid bias. *b*) Early ignition owing to v_g being in phase with v_a. The waveform of the critical grid voltage v_{cr} corresponding to v_a is indicated in broken lines. *c*) Corresponding anode current. *d*) Late ignition, owing to v_g lagging in phase with respect to v_a. *e*) Anode current corresponding to (*d*). t' is the moment of ignition.

[4]) See, e.g., D. M. Duinker, Relay valves as timing devices in seam-welding practice, Philips Techn. Rev. 1, 11-15, 1936, or J. W. G. Mulder and H. L. van der Horst, A controllable rectifier unit for 20,000 volts/18 amperes, Philips Techn. Rev. 1, 161-165, 1936.

[5]) Only approximately, because in the first place there is a certain minimum ignition voltage, in consequence of which the valve cannot strike exactly at the beginning of the half cycle, and in the second place owing to the voltage dropping below the arc voltage the valve extinguishes just before the end of the half-cycle. These points, however, are of no consequence here.

anode voltage, as is demonstrated in *fig. 3*. In the case considered here (a load consisting only of a resistance) the anode current assumes the shape of a cut half-sine.

In the foregoing only one half-cycle of the alternating voltage supply has been used. In order to make use of both valves two relay valves are connected in anti-parallel or back-to-back (*fig. 4*), with the two alternating grid voltages in anti-phase.

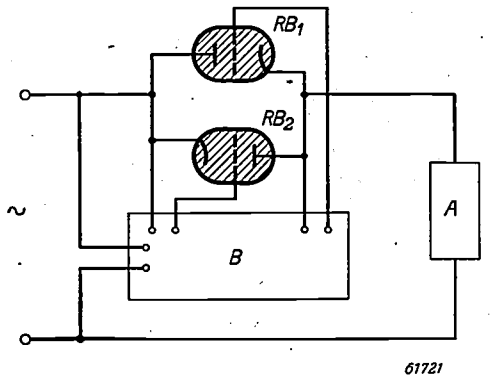


Fig. 4. Control device for alternating current with two relay valves (RB_1, RB_2) connected in anti-parallel. A is the apparatus controlled, B the system with which the moment of ignition of the valves can be adjusted.

Such control devices — which in essence are synchronous switches — have found an important application in electrical welding (both spot and seam welding), often for powers up to several hundreds of kVA. They may also certainly be of use for controlling the luminous intensity of incandescent lamps (with the advantage of easy remote control). But such a control system with the aid of relay valves particularly offers a good solution of the problem with TL lamps, for which no satisfactory method of control had hitherto been found. At first sight one would not expect this system of control with relay valves to be successful with TL lamps, for there is no doubt at all about the occurrence of zero current gaps in this case [6]). Furthermore, in one respect the control device with relay valves shows some analogy with the variable ratio transformer: if we call the quantity $\omega t' = \beta$ the ignition angle ($\omega = 2\pi$ times the mains frequency; t' = the instant of ignition) then in the part of the control range where the ignition angle lies between 90° and 180° the highest instantaneous value of the voltage at the lamp is generally less than the peak V_{max} of the mains voltage, namely $V_{max} \sin\beta$; if β approaches close enough to 180° this voltage ultimately becomes so small as to be insufficient for the ignition of the TL lamps.

On closer investigation, however, it is found that the position is not by any means so unfavourable as one might expect from the foregoing. When modern TL lamps are used and a simple measure is applied to which we shall revert presently, the working may even be said to be most satisfactory.

For an understanding of what is meant we have to explain what takes place in a system as illustrated diagrammatically in *fig. 5*. As soon as the grid voltage in one of the relay valves exceeds the critical value that valve ignites. Current then begins to flow through R_0, a resistor of the order of 10,000 ohms. The TL lamps (only one is drawn in the illustration) are connected in parallel to R_0, each in series with the usual choke. If the lamp and the choke were connected to the mains direct then we should have a (somewhat distorted) alternating current (fig. 1) lagging about 60° with respect to

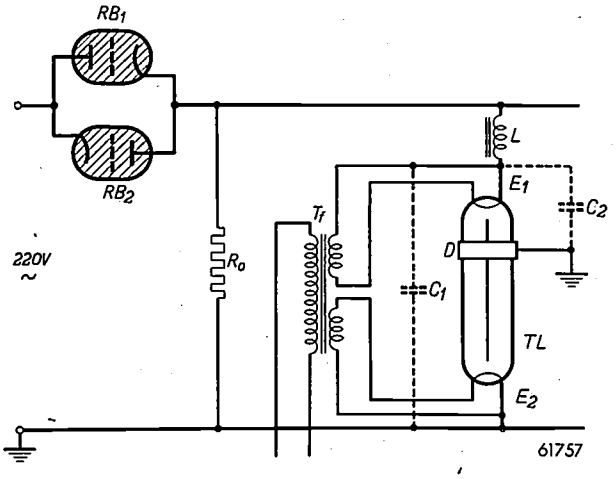


Fig. 5. Relay valves (RB_1, RB_2) used for controlling the luminous intensity of a TL lamp (TL). L = choke in series with each lamp, T_f = filament current transformer. In the middle of the lamp is the conducting strip along the glass tube, earthed via the clamping band D. For the meaning of R_0, C_1, C_2, E_1 and E_2 see the text.

the mains voltage. When, with the system according to fig. 5, the ignition angle β is adjusted to about 60° then the normal current flows through the lamp (the small voltage loss in the relay valves — the arc voltage — may be disregarded here). Upon the ignition angle being enlarged the current decreases, the extreme limit being reached at a value of about 1 mA per TL lamp. The ignition angle is then about 135°, so that at the moment of

[6]) This is not so evident as might appear from fig. 3. This diagram has been drawn for the case where the load is resistive. If there is also inductance in the circuit then under certain circumstances part of the control range may be free of zero current intervals. If the ignition is made to take place gradually later in the cycle then there is always a point beyond which a zero current interval occurs. In the case of a TL lamp with the normal choke the conditions are such that this point is reached very soon. To a good approximation it may therefore be said that a TL lamp controlled by means of relay valves always works with zero current gaps except at the full strength of current.

ignition the mains voltage still has an instantaneous value of about 220 V, which is not very far below the ignition voltage of a TL lamp with hot cathodes.

The fact that with $\beta = 135°$ the current is so much smaller than the nominal current is due to the presence of the inductance L and the arc voltage v_{arc}. The effect of these two quantities is that with increasing β the current falls much more quickly than it does in a circuit where there is only a resistance, as was assumed in the case of fig. 3.

Furthermore the voltage V_{max} in β, suddenly applied to the resistor R_0 each time one of the relay valves strikes, excites an oscillating circuit. This circuit is formed by the inductance L of the choke and the capacitance C_1 between the electrodes (and the filament-current windings connected thereto) of the TL lamp [7]). As a simple calculation shows, if the damping of this circuit and the stray capacitance of the choke were disregarded then the voltage across C_1 (thus across the TL lamp) would reach a peak value of about twice the amplitude of the mains voltage at the moment of ignition of the relay valve; thus with $\beta = 135°$ a peak value of about $2\ V_{max} \sin 135° = 440$ V. Owing to the causes mentioned, the actual value measured is not so high but still a good deal higher than $V_{max} \sin \beta$. Naturally this is all to the good for the ignition of the TL lamps.

Mention was made earlier of a modern design of TL lamps having a conducting strip. This strip is on the outside of the tube and runs in the longitudinal direction, extending almost to the electrodes. A potential difference between the electrodes is thereby mainly concentrated in the small spaces between each electrode and the end of the conducting strip. Consequently, compared with a lamp not having such a strip, the field strength at the electrodes is much greater and the ignition voltage of the lamp correspondingly lower.

The simple measure that has been referred to as contributing towards a considerable improvement — in the sense that the luminous flux can be further reduced without the lamps being caused to burn unsteadily or extinguish altogether — consists in earthing this conducting strip on the lamps (fig. 5). Between this strip and the non-earthed electrode E_1 we then have the capacitance C_2 instead of the capacitance C_1. The voltage peak is then no longer divided between two spaces but is available for one space (that at E_1). This facilitates ignition to such an extent that the smallest current on which the lamps can still burn steadily actually becomes in the order of 1 mA, thus making continuous control of the luminous intensity possible between the maximum and a very low level [8]). Measurements taken on normal TL lamps of 40 W, fed with a control device to be described later, showed that the lowest level at which the lamps still burn quite steadily is between 1/70 and 1/100 of the maximum level. This is so low that, if complete darkness is desired, one can then even interrupt the current entirely without any annoying drop in luminous intensity. It is to be noted that a large number of TL lamps can be controlled simultaneously with only one pair of relay valves; each lamp has to be connected in series with a separate choke in the usual way. Further, it is evident that no starters are needed. On the other hand, as already observed, filament current transformers are required in order to keep the cathodes at a sufficient temperature when the discharge current in the lamps is reduced; it is advisable to vary the heating current in the opposite sense to the discharge current, because with the full current plus the heating current the cathodes would become too hot.

There are various ways of bringing about this variation of the heating current. For instance, the filament current transformers can be fed from a small variable ratio transformer controlled by the same knob as that with which the device is operated for adjusting the ignition point of the relay valves. It may even suffice to vary the heating current step for step, for instance by arranging for contacts on the spindle of the control knob to switch on and off resistors connected in series with the filament current transformers.

A control device with relay valves is suitable not only for TL lamps but also for tubular fluorescent lamps working at high tension, in which case the device is connected in series with the primary of the high-tension transformer.

We shall first give a description of a control device built for 35 TL lamps, following this up with some details of the relay valves suitable for such an application.

Control device for 35 TL lamps

In *fig. 6* a photograph is given of the apparatus used for controlling the lighting of one of our demonstration rooms. This lighting consists of 35 TL lamps of 40 W. As the photograph shows, the apparatus comprises three parts: a casing in

[7]) See, e.g., Tj. Douma, Voltage impulses in rectifiers, Philips Techn. Rev. 9, 135-146, 1947. In this article a number of oscillograms are given showing oscillations such as are referred to here.

[8]) A provisional report on this method of control appeared under the title "Fluorescent light dimming control" in Philips Technical Communication, pp 9-13, 1948 (No. 7), published by Philips Electrical Industries of Australia Pty., Ltd., Sydney.

which the two relay valves, among other parts, are housed; a movable control box and a case containing a simple filter. The filter serves for preventing the oscillations mentioned above interfering with radio reception in the vicinity via the lighting mains.

The main elements of the circuit are represented in *fig. 7* .The manner in which the instant of ignition of the relay valves is varied by means of the variable

such a value that the valves could not strike if there were no other voltage in the grid circuit. These grid biases are supplied by two small auxiliary rectifiers, the valves of which (AZ 41) can be seen in fig. 6.

The other voltage referred to is an alternating voltage that is shifted in phase in order to change the moment of ignition. This voltage is not sinusoidal, as is represented in *fig. 8a*, but shows a steep

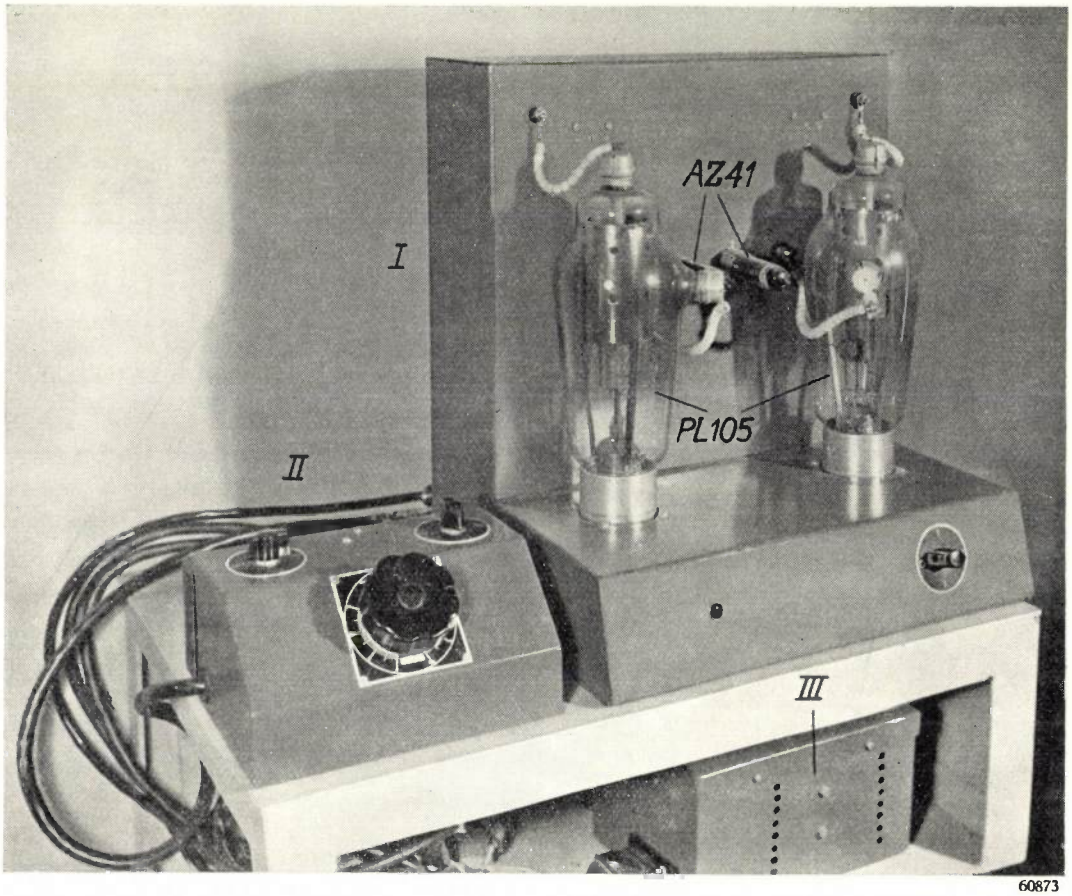


Fig. 6. Control device for 35 TL lamps of 40 W. *I* Opened casing with the two relay valves PL 105. *II* Movable control box. *III* Case with filter preventing radio interference via the lighting mains. The two rectifying valves AZ 41 supply the fixed negative grid bias for the relay valves.

resistor R contained in the control box will be explained later. The control box contains, further, two switches S_1 and S_2. If complete darkness is required then, after it has been reduced as far as possible, the current is interrupted with S_1. If the full light intensity is needed for some length of time then the control device can be put out of action by shorting it with S_2.

In this demonstration apparatus no provision has been made for varying the heating current.

The grid circuit calls for some explanation. Each grid receives in the first place a negative bias of

front, in this case in the shape of a peak (fig. 8*b*). Thus ignition always takes place at the moment of the voltage peak, regardless of any changes taking place in the characteristic representing the relation between the anode voltage and the critical grid voltage. In the case of a sinusoidal grid voltage (fig. 8*a*) these changes would cause undesired displacements of the moment of ignition.

The peak shape of curve with a steep front has been chosen because this can easily be produced, for instance, with the aid of a so-called peak transformer, i.e. a transformer with the secondary coil

mounted on a part of the core that becomes very highly saturated. *Fig. 9* shows how the core of such a peak transformer can be made. A sinusoidal alternating current passing through the primary coil induces in the said part of the core (K_2, fig. 9) used which reaches saturation at a much lower magnetic field strength than the conventional transformer sheet of which the rest (K_1) of the core is made. This material is an alloy of iron and nickel specially rolled and annealed [9]).

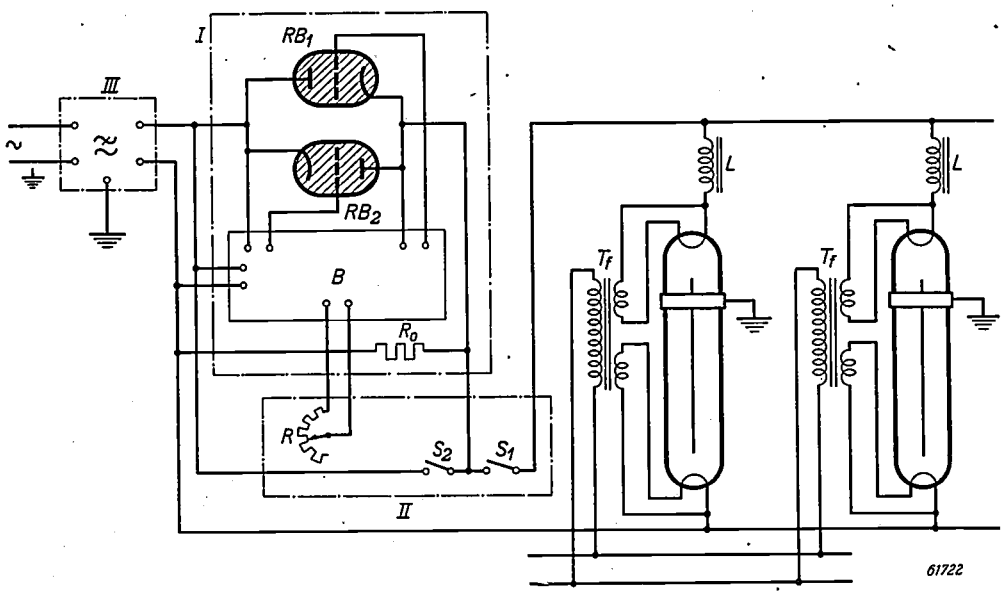


Fig. 7. Basic circuit showing the working principle of the control device illustrated in fig. 6. *I* Case containing the two relay valves (RB_1, RB_2) and the network B supplying the grid voltages. *II* Control box containing a small variable resistor R, with which the moment of ignition of the relay valves is varied, and two switches (S_1, S_2). L = chokes connected in series. T_f = filament-current transformers. *III* Mains filter.

a magnetic flux density B which, as function of the time t, has practically the shape of a "square sine" (*fig. 10*). Every time B changes in sign a voltage peak is induced in the secondary coil and the quicker B changes the steeper are the flanks of that peak. Since peaks with very steep flanks are desired in order to get a sharply defined moment of ignition, for the core part K_2 a material has been

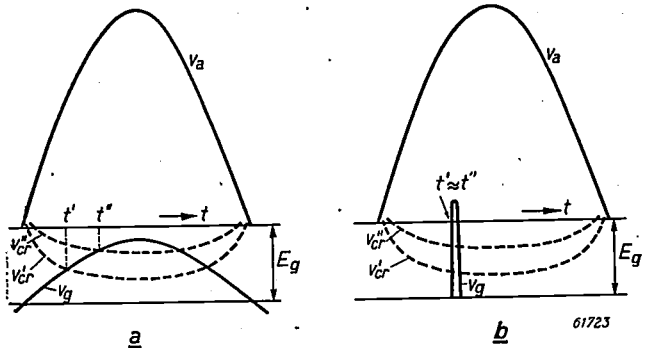


Fig. 8. Any variation in the characteristic of a relay valve, in the case of a sinusoidal alternating voltage on the grid (*a*), influences the instant of ignition ($t' \neq t''$). If, on the other hand, a positive peak-shaped grid voltage is used (*b*) a variation of the characteristic has no such influence and the moment of ignition remains practically constant ($t' \approx t''$).

For shifting the phase of the peaks the mains voltage is applied to a centre-tapped auto-transformer (*fig. 11a*), and to a fixed capacitor C and a variable resistor R (the resistor in the control box, fig. 7) connected in series. When R is increased from 0 to the maximum value R_{max} the phase difference α between the mains voltage and the voltage MP (M = centre of the coil AB, P = common point of C and R) increases from 0 to $\alpha_{max} = 2 \arctan \omega C R_{max}$ (see fig. 11*b*), where $\omega = 2\pi$ times the mains frequency; the voltage MP is constant in amplitude. When the peak transformer described above is connected between M and P the peaks can be shifted in phase by varying R; R_{max} has to be so chosen that the limits between which α can be varied (0 - α_{max}) coincide with the limits between which the peak has to be shifted.

In order to feed a peak voltage to the grids of the two relay valves either two peak transformers

[9]) J. L. Snoek, Magnetic cores for loading coils, Philips Techn. Rev. 2, 77-83, 1937, in particular fig. 4, curve *A*. The design of the peak transformer described is due to D. M. Duinker and J. L. Snoek of Philips Research Laboratory.

can be used or one such transformer with two separate secondary windings mounted on the saturated part of the core.

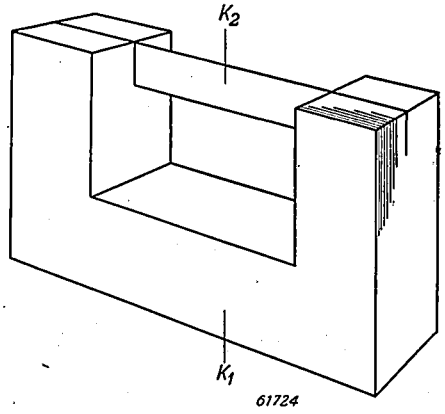


Fig. 9. Core of a peak transformer. K_1 part of the core made of the usual transformer sheet, carrying the primary windings. K_2 part of the core of a much smaller cross section and preferably made of an easily saturated material. The secondary coil is wound on K_2.

When current is taken off between P and M (fig. 11*a*) $\alpha_{max} < 2$ arc tan $\omega C R_{max}$, so that R_{max} has to be greater than (tan $\frac{1}{2}\alpha_{max}$)/ωC. Therefore, in order to keep R_{max} as small as possible, it is necessary to limit this current consumption to the utmost. Now the primary current of the peak transformer is for the greater part reactive, so that it can be compensated with a suitable parallel capacitor (C_3, *fig. 12*). In this way it has been possible to manage with a variable resistor of very small dimensions.

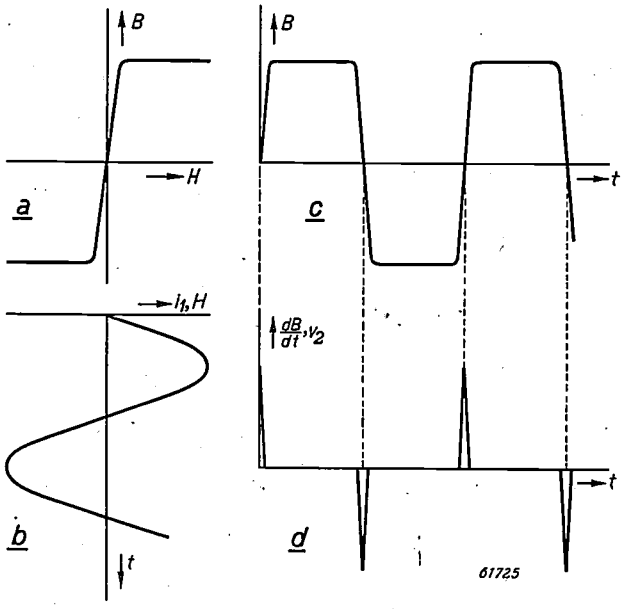


Fig. 10. Illustration of the operation of a peak transformer. *a*) Flux density B as a function of the magnetic field strength H for the part K_2 of the core (see fig. 9). *b*) H, proportional to the primary current i_1, as a function of time. *c*) B as a function of time. *d*) The voltage v_2, induced in the secondary coil, as a function of time. v_2 is proportional to dB/dt.

Another point to be noted is the following. The circuit is so arranged that the earliest ignition (maximum luminous intensity) takes place at $R \approx 0$. The corresponding moment of ignition is determined by the given amplitude of the mains voltage, the arc voltage and the re-ignition voltage of the TL lamps; it lies at about 60° beyond the zero point of the mains voltage. Now, as fig. 10 shows, the peaks occur when the primary current of the peak transformer passes through zero, so that at $R \approx 0$ this current must have a very definite phase shift with respect to the voltage of the mains from which the TL lamps and the coil AB (fig. 11*a*) are fed. This fixed phase shift has been provided for by connecting a capacitor (C_4, fig. 12) in series with the primary of the peak transformer.

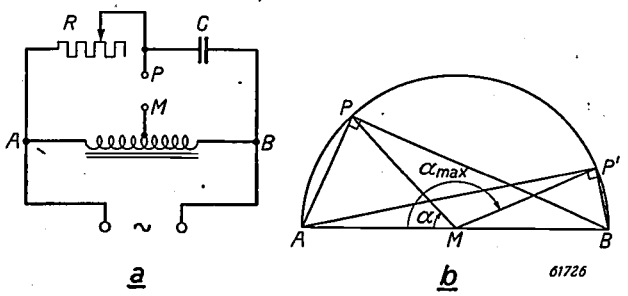


Fig. 11. *a*) Variation the resistor R causes a change in the phase displacement α between the voltage MP and the mains voltage AB, while MP remains constant in amplitude (M is the centre of AB), as shown by the vector diagram (*b*). AP corresponds to the voltage across R, BP to that across the fixed capacitor C. At the maximum value of R the point P comes to lie at P', corresponding to the maximum phase shift α_{max}.

The number of TL lamps that can be connected to one pair of relay valves is limited by the maximum permissible mean current for the valves. For the PL 105 type of valve used in the control device described here this current is 6.4 A (per valve). When burning at full strength each 40 W TL lamp demands about 185 mA from the mean current of each of the relay valves, so that with two PL 105 valves at most 6.4/0.185 = 35 TL lamps of 40 W can be used.

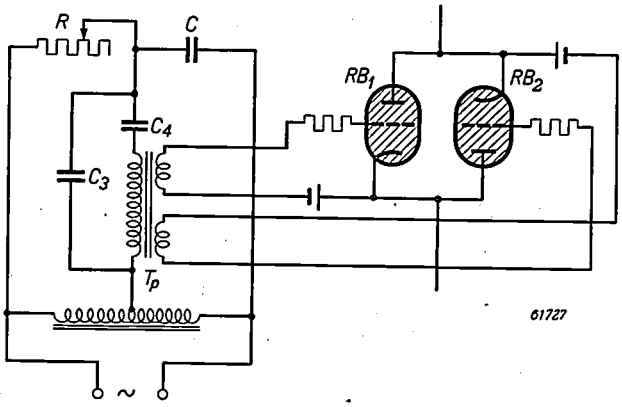


Fig. 12. Details of the phase-shifting and peak-voltage circuits. T_p = peak transformer with two secondary windings supplying a peak voltage to the two relay valves (RB_1, RB_2). R and C as in fig. 11*a*. C_3 capacitor for compensating the reactive current of the circuit C_4-T_p. C_4 capacitor for giving the peaks the right phase, at $R \approx 0$, corresponding to the full luminous intensity.

Usually the current rating of a TL lamp is expressed by the R.M.S. value, and for the 40 W type this is 420 mA. To derive from this figure the required demand from the mean current of the relay valves we assume that the control device is connected to only one TL lamp burning at full strength and presume, for the time being, that the current flowing through the lamp is purely sinusoidal and thus has a form factor $\pi/2\sqrt{2} = 1.11$. Taking this current to be commutated, its mean value would therefore be 420/1.11 = 378 mA. Each of the relay valves allows only a half-cycle of this current to pass at a time and therefore its current has a mean value of $\frac{1}{2} \times 378 = 189$ mA. Owing to the deviation from the sinusoidal form (see fig. 1*b*) the form factor is slightly larger and thus the mean current slightly less (185 mA).

In large halls, where usually far more than 35 lamps are needed, it is best to divide the lamps into three groups of at most 35 and to feed each group from a different phase of the mains. Each group is fed via one pair of relay valves; if desired, these three pairs can be controlled simultaneously with only one knob.

Some details of the relay valve PL 105 and similar types

Since the PL 105 type of relay valve used in this control device has never been described in this journal, this opportunity is taken to say something about it.

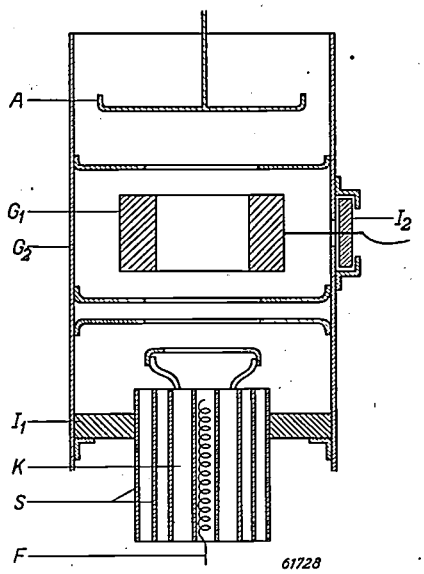


Fig. 13. Cross section of the electrode system of the relay valve type PL 105. K = cathode indirectly heated by the filament F. S = screens for heat insulation. G_1 = control grid, G_2 = screen grid, A = anode. I_1 insulating plate fixing the screen grid with respect to the cathode. I_2 one of the three insulators by means of which the control grid is fixed to the screen grid.

The PL 105 valve has four electrodes (see the cross-sectional drawing in *fig. 13*): an indirectly heated oxide-coated cathode, a control grid, a screen grid and an anode. The gas filling consists of saturated mercury vapour.

The cathode is formed by two concentrical cylinders connected by some radial partitions. The heating filament is inside the smaller cylinder. The surfaces taking part in the emission are the inside of the large cylinder, the outside of the small one and the two sides of the partitions. Enveloping the larger cylinder are two non-emitting cylinders acting as heat insulators so as to minimize the filament current required to keep the cathode at the right temperature. The filament, the cathode and the screen grid are connected to the pins of the valve base via a glass pinch.

The control grid is a graphite ring connected by a flexible wire to a chrome-iron contact cap fused air-tight onto the side of the glass bulb.

The anode is a metal disc with its edge rounded off so as to avoid excessive field strengths. It is connected to a terminal cap (likewise of chrome-iron) on top of the bulb.

The screen grid consists of metal walls with a central opening, mounted between the cathode and the control grid and between the control grid and the anode, and further a metal can enveloping all the other electrodes. The purpose of the screen grid is to shut out all external stray fields and to keep the capacitance between anode and control grid as small as possible with respect to the capacitance between control grid and cathode. The object of the latter is to prevent a sudden change in the voltage on the anode causing the control grid vol-

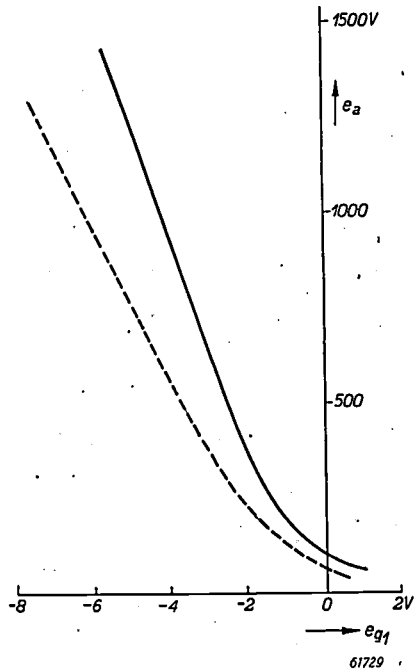


Fig. 14. Average characteristic of the relay valve PL 105; e_a = anode voltage, e_{g1} = control-grid voltage at which the valve just strikes. Fully-drawn line: with screen-grid voltage zero. Broken line: with a certain, positive, screen-grid voltage.

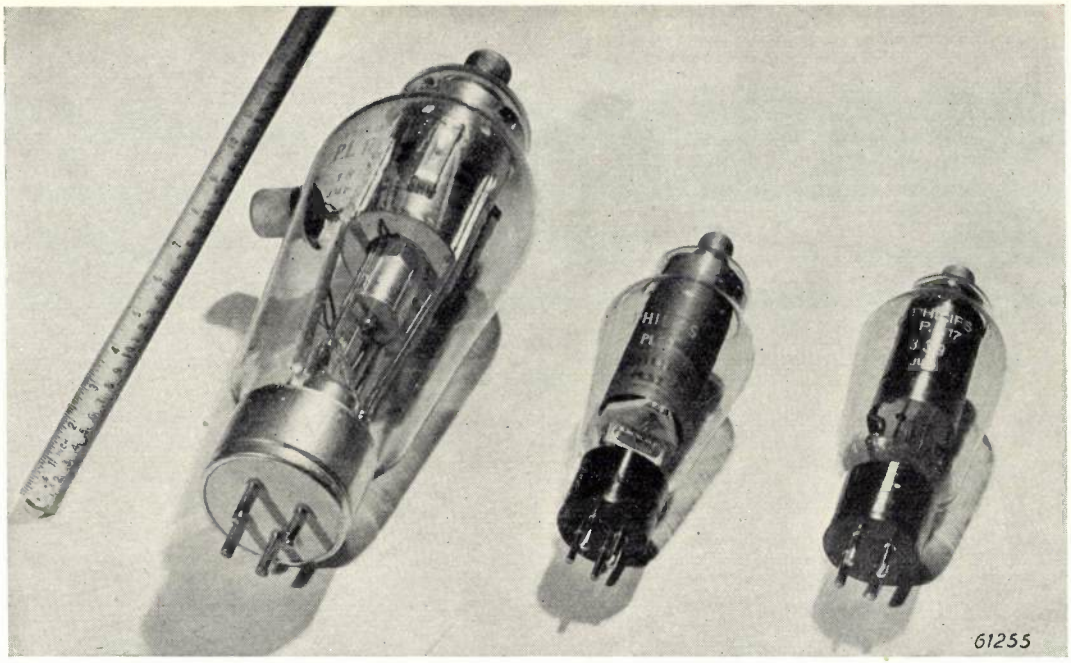


Fig. 15. Three relay valves of similar construction. From left to right: PL 105 (maximum permissible mean anode current 6.4 A), PL 57 (2.5 A) and PL 17 (0.5 A).

tage to exceed momentarily the critical value and thus prematurely igniting the valve. For this purpose the screen grid has to be connected to the cathode across a not too high impedance, say a resistance of 10,000 ohms. For certain applications it is also possible to apply a voltage between the control grid and the cathode: if, for instance, this grid is made positive then with a given control-grid voltage a lower anode voltage is sufficient to cause the valve to strike. The characteristic representing the anode voltage at which the valve strikes, as a function of the control-grid voltage and with a constant screen-grid voltage (*fig. 14*), is thereby shifted to the left. In this way small variations of the characteristic can be corrected.

The characteristic is strongly dependent upon the position of the two grids with respect to each other and with respect to the cathode. These electrodes must therefore be rigidly fixed. The screen grid is fixed with respect to the cathode by the ceramic bottom of the screen-grid can. This can carries three rod-shaped insulators, likewise of a ceramic material, in which the control grid is mounted by means of three radial rods.

The maximum permissible mean current, as already stated, is 6.4 A. Another criterion that may be decisive in certain cases is the peak current, which may not exceed 40 A.

The peak value of the voltage between cathode and anode of the PL 105 valve must not exceed 2500 V, both in the positive and in the negative sense. Otherwise short-circuiting is apt to occur, either — with positive anode — in the forward direction (but at an undesired moment beyond the control of the grid) or in the inverse direction (backfiring), which is always undesirable and, moreover, may be detrimental to the valve.

Two smaller types (*fig. 15*), the PL 57 for a maximum mean current of 2.5 A and the PL 17 for 0.5 A max. mean current, have been developed on mainly the same lines as the PL 105 described here. The PL 57 has an indirectly heated cathode, whereas the PL 17 has a directly heated cathode. These small valves have no screen grid.

Summary. The usual methods for controlling the luminous intensity of incandescent lamps, for instance with the aid of a variable ratio transformer or a series resistor, do not lend themselves for application with TL or other gas-discharge lamps. A good control device for such lamps (on A.C. mains) consists of a pair of relay valves connected in anti-parallel, the ignition point of which can be varied, for instance, by means of a grid voltage that is shifted in phase. A perfectly gradual variation of the luminous intensity of TL lamps between the maximum and a very low value (70 to 100:1) is possible when the conducting strip provided on modern lamps is earthed. To keep the cathodes of the lamps at the right temperature it is necessary that they be heated with a filament current so long as there is no discharge current. For this purpose separate transformers are needed; on the other hand starters can be dispensed with. With this method of control, which lends itself well for remote control, TL lamps can be used also in theatres, cinemas and lecture halls, for stage lighting and other purposes where the luminous intensity has to be gradually lowered or raised. The same applies for high-tension fluorescent lamps. A demonstration apparatus serving 35 TL lamps of 40 W is described. Here the alternating grid voltage is peak-shaped; it is derived from a special transformer and can be shifted in phase by means of a small variable resistor. Some particulars are given of the relay valves employed.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the address on the back cover.

1904: J. M. Stevels: Relations entre les structures des verres et leurs propriétés mécaniques, physiques et chimiques (Verres et Refractaires **3**, 359-368, Dec. 1949). (Relations between the structures of glasses and their mechanical, physical and chemical properties; in French.)

Zachariasen's theory (see No. **1774** of these abstracts and Philips Techn. Rev. **8**, 231-236, 1946) concerning network-forming and network-modifying ions in glass and its experimental confirmation by the work of Warren c.s. is treated first. After having discussed the structure of borate glasses (existence of an "accumulation region" and a "destruction region"), the writer indicates certain improvements of this theory. The same ion may, according to circumstances, behave either as a network-forming or a network-modifying iron. According to Smekal the glass network is not only due to an irregular arrangement of oxygen polyhedra but it also depends upon the nature of the bonds within the polyhedra. These bonds can be either heteropolar or homopolar, thereby influencing the distance between the ions involved. The influence of these new conceptions on the interpretation of mechanical, physical and chemical properties of glass is indicated. An important constant, determining the rigidity of the network, is $Y = 2Z - 2R$, where Z is the average number of oxygen ions surrounding a network-forming ion and R is the ratio of oxygen ions to network-forming ions. Y equals the number of bridging oxygen ions per polyhedron. For commercial glasses $4 > Y > 3$, for soft glasses $3 > Y > 2$, whilst glasses with $Y < 2$ show a strong devitrifying tendency. The more rigid the lattice, the more the thermal expansion decreases, whereas the melting point increases.

Finally conclusions may be drawn from the electrical properties of glass as regard its structure and especially as regards the tendency of certain ions to change over from a network-modifying position into a network-forming position.

1905: R. van der Veen: Induction phenomena in photosynthesis, II (Physiologica plantarum **2**, 287-296, 1949).

In this paper (see No. **1882** of these abstracts) induction phenomena experiments on secondary peaks of the adaption curve are discussed. These experiments make it probable that photosynthesis is inhibited by an agent which is formed during photosynthesis and acts with a certain lag of time. The secondary peaks have their origin in this slow action of the inhibitor, which may be O_2 or perhaps some organic peroxide.

Experiments in nitrogen with very little oxygen show that the inactivation of the adaption factor is probably caused by an oxidation of this factor, while its activation may be caused by a reduction by the illuminated chlorophyll complex.

Experiments in pure nitrogen with CO_2 make it probable that the initial CO_2-uptake is damaged in dark under anaerobic conditions, but that recovery takes place when oxygen is present.

Experiments in hydrogen confirm the supposition, made in the first paper, that by heat-treatment of leaves (exposing leaves of Holcus lanatus to 48° C during 3 minutes) the connection between H-donor and chlorophyll is damaged.

1906: H. C. Hamaker: Systematische en toevallige fouten bij het aflezen van de stand van een wijzer op een schaal (Statistica **3**, 209-223, 1949, No. 5/6). (Systematic and random errors in estimating the position of a pointer on a scale; in Dutch.)

Using data published by V. Varangot the writer analyses the systematic and random errors made in estimating the position of a pointer on a scale. At positions between 0.1 and 0.5 and between 0.9 and 1.0 there is a tendency to estimate the position a few 0.01 too low, whereas at positions between 0 and 0.1 and between 0.5 and 0.9 the reverse tendency is evident. In a number of cases the systematic and the random errors are separated by a regression analysis. Personal differences in the magnitude of the systematic errors are investigated. (See No. **1817** of these abstracts.)

1907: F. A. Kröger and W. Hoogenstraaten: The location of dissipative transitions in luminescent systems (Physica The Hague **16**, 30-32, 1950, No. 1).

From the temperature-dependence of the decay of fluorescence in $Cd_2B_2O_5$-Mn and $CdSiO_3$-Mn it is found that the quenching of fluorescence at high temperatures is due to dissipative transitions from the excited state of the centres, whereas the quenching at low temperatures is due to dissipative transitions from higher states of a different character. For the orange samarium fluorescence of $CaWO_4$-Sm the situation is similar, the only difference being that at low temperatures the energy is not dissipated but emitted as tungstate fluorescence.

1908: H. G. Beljers: A demonstration of the induced magnetic moment in the third direction at gyromagnetic resonance (Physica, The Hague **16**, 75-76, 1950, No. 1).

If a material showing gyromagnetic resonance is placed in a constant magnetic field on which an alternating field is superimposed perpendicularly, an alternating magnetic polarization occurs not only in the direction of this alternating field but also in the third direction, perpendicular to both the alternating and the constant field. The existence of this polarisation is proved, using a system of wave-guides (magic tee). (See Philips Techn. Rev. **11**, 313-322, 1950, No. 11, and No. **1850** of these abstracts).

1909: C. J. Bouwkamp: On the freely vibrating circular disk and the diffraction of circular disks and apertures (Physica, The Hague **16**, 1-16, 1950, No. 1).

The writer develops a theory of the acoustic field produced by a freely vibrating, rigid, circular disk on the assumption that the wavelength is large compared to the radius of the disk. The solution is presented in the form of a series of ascending powers of wave number times radius of disk. The new approach, which is based on integral equations, easily permits the explicit calculation of a number of terms of this series. The results are equally applicable to the diffraction at circular disks and apertures of plane scalar waves impinging in the normal direction upon the obstacle. A survey of earlier results by various authors is included.

1910: R. Loosjes and H. J. Vink: Distribution du potentiel dans la couche d'une cathode à oxydes pendant une impulsion de courant de grande densité (Le Vide **5**, 731-738, Jan. 1950). (Potential distribution in the oxide layer of an oxide-coated cathode during a current impulse of high density; in French.)

The distribution of the potential in the coating of an oxide-coated cathode is investigated according to two different methods. Firstly the distance of the anode to the cathode is varied and the measured potential differences are extrapolated to zero distance. In this way the value of the potential at the surface of the coating is found. Secondly the potential distribution in the coating is determined by a probe method. No jump in potential is found at the interface between metal and oxide. The greater part of the potential drop is concentrated near the outer surface of the coating.

1911: W. Nijenhuis: Theoretische grenzen voor de overdracht van brede frequentiebanden (T. Ned. Radiogenootsch. **15**, 13-31, 1950, No. 1). (Theoretical limits for the transmission of broad frequency bands; in Dutch.)

The paper deals with the question of obtaining maximum gain in a given frequency band from two-terminal and four-terminal interstages. It has been attempted to elucidate and, partly, to generalize the treatment in Bode's: Network Analysis and Freedback Amplifier Design.

1912*: E. W. Gorter: L'aimantation spontanée de ferrites ferromagnetiques à structure de spinelle (C.R. Acad. Sc., Paris **230**, 192-194, 1950, Jan. 9). (Spontaneous magnetisation of ferromagnetic ferrites with spinel structure; in French.)

The spontaneous magnetisation of a number of ferrites of the type MFe_2O_4 (M = bivalent metal) and of mixed crystals of these ferrites with zinc ferrite ($ZnFe_2O_4$) is measured; the results confirm Néel's theory (Ann. Physique **3**, 137, 1948).

1913: J. L. Meijering and G. W. Rathenau: Rapid oxidation of metals and alloys in the presence of molybdenum trioxide (Nature, London **165**, 240, 1950, Febr. 11).

The influence of molybdenum trioxide on the oxidation of metals and alloys has been studied as a function of temperature. A graph is given of the depth of penetration in 8/92 Al-Cu wires for different times of heating ($^1/_2$ h, 3 h, 20 h, 165 h) in the temperature range from 400 °C to 550 °C.

The results of these and other experiments support the view that the occurrence of a liquid oxide phase is the determining factor, not the dissociation of MoO_3 as put forward by Leslie and Fontana.

1914: E. J. W. Verwey: Atomic arrangement in spinels in connection with their physical properties (Proc. 11th Int. Congr. Pure and Appl. Chem., Vol. **I**, 285-291, Febr. 1950).

Survey of investigations regarding the structure and the electrical and magnetic properties of spinels (see Nos **1738**, **1739**, **1845** of these abstracts and Philips Techn. Rev. **9**, 185-190, 239-248, 1947).

1915: K. F. Niessen: On avoiding low frequencies in a rectangular cavity resonator used as part of a triode generator (Appl. sci. Res., The Hague **B1**, 325-340, 1950).

Reprint of an article published on the occasion of the Marconi Congress, Rome 1947 (Roma, Consiglio Naz. delle Ricerche 1948, 312-329). It is shown that by adding a "side room" to a prismatic cavity of quadratic or triangular cross-section (see Nos **1815**, **1853**, **1854** of these abstracts) all lower frequencies except one (the frequency required) are changed. By combining such extended cavities, prismatic resonators are obtained with a symmetrical cross section, in which vibrations with undesired frequencies may be suppressed by applying a bar at a suitable place.

1916: E. J. W. Verwey: Theory of the electric double layer of stabilized emulsions (Proc. Kon. Ned. Akad. Wetensch. Amsterdam **53**, 376-385, 1950, No. 3).

The electrical potential function and the distribution of the charges at the interface of two immiscible liquids are calculated with the aid of a Gouy-Chapman type of theory, for the case that in addition to the original double layer a surface charge is present at the interface.

1917: K. F. Niessen: On one of Heisenberg's hypotheses in the theory of specific heat of superconductors (Physica, The Hague **16**, 77-83, 1950, No. 2).

An assumption made by Heisenberg in Koppe's theory of the specific heat of superconductors is discussed. A comparison with another problem seems at first sight to contradict this assumption but turns out to support it. Heisenberg's assumption is also confirmed in a direct way.

1918: K. F. Niessen: The energy of the normal electrons in a superconductor as a function of temperature and thickness of the superconducting layer on the Fermi surface (Physica, The Hague **16**, 84-94, 1950, No. 2).

In order to simplify calculations Heisenberg's assumption about the superconducting layer covering a part of the Fermi surface (see No. **1917**) is modified in so far that a layer is assumed of thickness Δ (by which is meant the difference of the energies corresponding to the upper and lower surfaces of the layer), which layer, covering a part ω of the Fermi surface, is entirely inaccessible to normal electrons, whilst all its volume elements h^3 are supposed to be occupied by superconducting electrons. A formula for the energy of the normal electrons is derived, which is a function of ω, T and Δ and which for $\Delta \gg kT$ coincides with Koppe's expression in the Heisenberg theory but for $\Delta \ll kT$ is identical with the value well known from Sommerfeld's theory of metals. A relation is derived between the transition temperature and the numbers of superconducting and normal electrons per cubic centimeter, both at $T = 0$.

1919: J. L. H. Jonker: Buizen met lintvormige electronenbundel; contact-, schakel-, kiesen telbuis (T. Ned. Radiog. **15**, 37-52, 1950, No. 2). (Valves with ribbon-shaped electron beam; contact valve, switch valve, selector valve, counting valve; in Dutch.)

Paper covering the subject already referred to under No. **R 126.**

1920: K. S. Knol: Electromagnetische golven in rechthoekige golfpijpen (T. Ned. Radiog. **15**, 53-74, 1950, No. 2). (Electromagnetic waves in rectangular wave guides; in Dutch.)

A survey of wave guide theory is given. The method of Brillouin for rectangular wave guides is dealt with in more detail and the solution of Maxwell's equations in the interior of a wave guide is found by superimposing the incident wave upon three reflected waves occurring when a plane linearly polarized electromagnetic wave strikes against two mutual perpendicular metal walls. Some applications of wave guides in practice are mentioned. A rubber sheet model may be used in studying problems of propagation of TE_{10} waves in rectangular wave guides. (See Philips Techn. Rev. **11**, 156-163, 1949, No. 5.)

VOL. 12 No. 4, pp. 97-128 OCTOBER 1950

Philips Technical Review

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

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

WIRE AND CABLE INSULATING MATERIALS WITH POLYVINYL CHLORIDE AS BASE

by E. A. J. MOL *).

621.315.336.96:678.773.312

The non-inflammability and much greater durability of certain plastic materials, especially those based on polyvinyl chloride, are the main reasons why these are being preferred more and more, over rubber and textiles, for insulating electrical conductors. Connecting leads insulated with these plastics are already being used on a large scale for radio and other electrical apparatus. In several countries, among which the U.S.A., large quantities of wire insulated with plastic material are also being used for permanent electrical installations in houses. In some other countries the application of these new materials is hindered by the insulation resistance requirements of official testing specifications. These requirements are based on the properties of rubber; polyvinyl chloride plastics have a lower insulation resistance, which, however, is quite high enough for almost all practical purposes. It would be regretted if the good properties of these new insulating materials could not be utilized to best advantage merely on account of the lack of suitable testing specifications.

Introduction

As insulating material for electrical conductors, rubber — either alone or in combination with a textile, such as cotton — still holds a predominating position. For many years it could indeed lay claim to being the most suitable material for this purpose: it has a very high insulation resistance, is easily processed and is not costly. Certain unfavourable properties, such as its inflammability and its lack of resistance to high temperatures and ultra-violet rays, have had to be accepted. Its inflammability, in particular, is a serious limitation: faults in electrical wiring rank second on the list of the causes of fires breaking out in houses ! It is not surprising, therefore, that materials possessing better properties than rubber should have been sought, and it is even less surprising that they should have been sought in the first place among plastics, of which in the last 20 years an enormous number of different kinds with a great variety of properties have been developed. This development, it may be remarked, has been stimulated more by the threat of war in the thirties, which made it necessary to reckon with a severe shortage of rubber in all industrial countries, rather than by the desire for a better material.

As a consequence a number of plastic insulating materials have been developed which are characterised by the common feature that they are all based upon the synthetic resin polyvinyl chloride as main ingredient, with certain admixtures designated according to their effect as plasticizers, stabilizers, pigments, etc., which will be dealt with farther on. The properties of these PVC materials [1]) depend very largely upon the nature and quantities of these admixtures, and it is possible to produce, even within this limited range of plastics, a series of materials each combining certain favourable properties.

*) Pope's Wire and Lamp Works, Venlo, Holland.

[1]) The abbreviation PVC is used in this article to denote the insulating materials based on polyvinyl chloride (thus including the admixtures); where the word polyvinyl chloride is written in full this relates to the synthetic resin itself.

In this way insulating materials have been manufactured which in many respects are far superior to rubber and are therefore being used more and more in preference to rubber. Nevertheless many are still averse to using PVC materials, mainly for one of the following reasons:

1) PVC is thought to be only a substitute for rubber, and as such only to have served its purpose under the exigencies of war.

2) It is considered that by no means sufficient experience has been obtained with PVC as insulating material, so that its use on a large scale would be too risky.

The first belief may be explained by the fact that during the war (and even for some time after it) there was a great scarcity of good plasticizers. The poor reputation of many of the PVC materials manufactured during that period (brittleness, stickiness and poor insulation) apparently still persists. At the present time, however, there is ample production of good plasticizers and results are now being obtained which ten years ago were considered to be impossible. Any fear of inferior quality is therefore now quite unjustified.

As regards the second point, i.e. the supposed lack of experience, extensive investigations were carried out both in America and in Germany even before 1940. As an example may be mentioned a test carried out in America with a PVC-insulated cable laid, without any other protection, 30 cm deep in the ground, and carrying continuously a voltage of 115 V. After twelve years the properties of the PVC material were still such that the cable could have rendered many more years' service. Data collected in Germany about PVC materials come mainly from the postal authorities, who report satisfactory results both with underground cables (laid in 1935) and with PVC wires used in telephone exchanges, notwithstanding the fact that the PVC materials used there were not so good as those used in America.

Experience on a very much wider scale has been gained during the war. Results with PVC on ships of the U.S.Navy, where particularly strict non-inflammability requirements had to be met, were so satisfactory that it was decided to use wire insulated with this material also for the permanent electric wiring of houses. The fact that it stands up against mechanical wear, strong sunlight and moisture has been particularly demonstrated by the use of PVC for field telephone lines run out along the ground, often across rough country and in a tropical climate.

Prejudice against PVC materials is in some cases also due to the lack of understanding that any material, whatever it may be, has to be used in a manner appropriate to its properties. It is, for example, often overlooked that PVC is thermoplastic, that is to say it becomes plastic above a certain temperature, so that any mechanical pressure accompanied by too high a temperature may cause short-circuiting. If simple precautions are taken this danger can usually be avoided, but if through ignorance or negligence it is omitted to take such measures then there is a chance of PVC unjustifiably getting a bad name.

The purpose of this article is to give more publicity to the properties and characteristic features of PVC materials, and in particular those of the PVC products being manufactured by Pope's Wire and Lamp Works in Holland since before 1940 and marketed under the trade name "Podur".

The composition of PVC materials

As already stated, for the manufacture of insulating materials with polyvinyl chloride as a base a number of substances are added, each for a particular function, which will be explained below. Roughly speaking, the composition of a PVC mixture is:

	parts by weight
1. polyvinyl chloride	100
2. plasticizer(s)	30 - 80
3. stabilizers	1 - 10
4. pigments	0.1 - 2
5. fillers	0 - 20
6. lubricants	0.5 - 2

Polyvinyl chloride

Pure polyvinyl chloride is a white powder obtained by the polymerization [2]) of vinyl chloride, $CH_2 = CHCl$. There are various methods of producing vinyl chloride. In the Netherlands a method is applied starting from ethylene (obtained from the mineral oil or coal industries) and chlorine (from the salt industry), with the formation of 1,2-dichloro ethane in an intermediate stage:

$$\underset{\text{ethylene}}{CH_2 = CH_2} + \underset{\text{chlorine}}{Cl_2} \rightarrow \underset{\text{1,2-dichloro ethane}}{CH_2Cl - CH_2Cl},$$

$$\underset{\text{1,2 dichloro ethane}}{CH_2Cl - C_2HCl} \rightarrow \underset{\text{vinyl chloride}}{CH_2 = CHCl} + \underset{\text{hydrochloric acid}}{HCl}.$$

[2]) See, e.g., J. C. Derksen and M. Stel, Plastics and their application in the electrotechnical industry, Philips Techn. Rev. 11, pp 33-41, 1949 (No. 2), in particular p. 35.

In the second stage of this process pyrolysis is applied (decomposition by heating).

The vinyl chloride obtained, which under atmospheric pressure and at room temperature is a gas but in this process is liquefied under pressure, is purified and then polymerized, when numerous monomer molecules (molecular weight 62.5) link up into long threadlike molecules with a molecular weight between 10,000 and 20,000. This means that 150 to 300 vinyl chloride molecules combine. A fully stretched molecule of polyvinyl chloride may reach a length of 0.08 μ.

This polymerization, which takes place very readily, is accelerated by catalysts (promoters), for instance benzoyl peroxide or hydrogen peroxide. It is exothermic, the large amount of heat released by the reaction (1300 calories per gram molecule of vinyl chloride) having to be carried off, because the temperature has to be kept within narrow limits, in order to obtain a constant product. Technically this can only be done properly by polymerizing in an aqueous emulsion, each dispersed droplet of vinyl chloride being surrounded by a quantity of water acting as a cooling agent. Moreover this method has the advantage that the polyvinyl chloride can thereby be obtained in a form resembling rubber latex, which lends itself well to further processing. Polyvinyl chloride latex is easily coagulated with certain agents, and, properly applied, this method results in the production of polyvinyl chloride in the form of a powder, which is easy to wash and dry. The latex can also be processed as such.

Plasticizers

Polyvinyl chloride obtained in the manner described in the form of a white powder is thermoplastic. By a combination of heat and pressure it can be formed into a horny substance, but there are very few uses for which this is suitable. When it is desired to manufacture materials resembling rubber a concentrated solution has to be made of polyvinyl chloride in non-volatile, usually liquid, solvents, which are called plasticizers or softeners.

The preparation of the solution, which is actually not liquid but rather like a gel, is called gelatinizing. This takes place at an elevated temperature (150 to 170 °C), either between pairs of rolls rotating at different speeds or in a powerful kneading machine followed by sheeting out. In both cases the gelatinized mixture is produced in the form of a sheet of 2 to 4 mm thickness, which is then either cut up into chips in a rotating cutting machine or formed into strands in an extruding machine. When the material is to be used for insulating wires the chips or strands are fed into the wire-coating machines, about which more will be said later; first we have to deal further with the plasticizers.

It has already been intimated that the quality of a PVC product stands or falls with the quality of the plasticizers. These substances are therefore of just as much importance as the polyvinyl chloride itself. They have to answer a number of requirements, some of which are opposed to each other. One of the most important conditions has already been mentioned: a plasticizer must not be volatile, otherwise it would gradually evaporate from the PVC material, which would then become hard and brittle, especially at high temperatures. This has been the experience with certain kinds of PVC products during the war, in which the rather volatile dibutyl phthalate was used as a plasticizer. Nowadays sufficient supplies of plasticizers are available which do not show this fault at all, as for instance tricresyl phosphate.

Apart from evaporation, an inferior plasticizer may leave the PVC material in various other ways and thus equally detract from the durability of such a material. For example, some are subject to decomposition (caused by high temperature or by high relative humidity such as occur in the tropics, and also from the action of ultra-violet rays), or else through migration they may be absorbed in some other plastic or oil with which the PVC may come into contact.

This latter phenomenon is governed by the mobility of the plasticizer and by the extent to which the contacting plastic has an affinity for it. This will be reverted to later.

Stabilizers

Polyvinyl chloride contains more than 50 wt % of chlorine, so that it is not surprising that at high temperatures HCl tends to become separated. This is an autocatalytic process, the HCl formed promoting the separation of still more HCl. It is therefore necessary to counteract this separation right from the beginning, and this can be done by adding a substance (the stabilizer) as for instance lead stearate, which binds HCl in an inactive form. With such a stabilizer the PVC material can withstand temperatures up to 150 °C for a long time without any adverse consequences.

Other substances serve as stabilizers for counteracting the effect of ultra-violet rays. Both the plasticizer and the polyvinyl chloride itself can be made insusceptible to these rays by means of suitable stabilizers. The scope of the present article does not allow us to go further into the mechanism of this process.

Pigments

The fact that PVC materials can be made in all sorts of striking colours and combinations of colours is an important advantage. In the case of multiple-core cables which form the nerves of a telephone exchange, and also for radio and similar apparatus, it is of great value to be able to distinguish readily the various electrical circuits from the colours given to the leads.

At first the pigments employed were those used for colouring rubber, but now preference is given to specially prepared pigments of greater dispersability, which at the same time stand up better against heat and light.

By a certain method of extrusion (the process described below by means of which the PVC material is applied to the wire that is to be insulated) or of finishing, a wire can be marked with a second colour, thereby considerably extending the number of circuits that can be distinguished.

Fillers

The insulating capacity and the resistance to heat of the PVC material are improved by the addition of small quantities of certain kinds of clay. Such fillers also have the advantage of reducing cost owing to their cheapness, but the manufacturer should not be tempted into using excessive amounts because then the final product would suffer considerably in quality. The same is the case when other fillers are used, so this is a point that has to be carefully watched.

Lubricants

The last of the admixtures to be mentioned are the lubricants, which are added so as to facilitate extrusion. Their action is based on the fact that they are not readily miscible with the other ingredients. They form a thin lubricating layer on the outside of the mass and thus prevent direct contact between the mass and the hot metal wall of the extruding machine. Incidentally this layer also reduces the risk of decomposition and discoloration and at the same time gives a higher gloss to the extruded material.

Among the substances used as lubricants are the salts of fatty acids (magnesium stearate, calcium stearate) and, sometimes, also paraffin wax, in quantities of 0.5 to 2%.

Properties of PVC materials

After this brief review of the constituents of PVC mixtures we have to consider the properties of PVC materials and how these are affected by the nature and quantities of the admixtures. Speaking in general terms it may be said that all the admixtures affect the quality of the final product; this applies especially to the plasticizers, in some cases also to the pigments, and even to a certain extent to the agents used in the preparation of the polyvinyl chloride itself (emulsifiers, catalysts, coagulators). On the one hand the multiplicity of the effects of these admixtures is a favourable factor, since it affords the possibility of manufacturing a great variety of products, but on the other hand it is obvious that for the manufacture of a product possessing specific properties a thorough investigation and close control over all stages of the manufacturing process are essential.

We shall classify the properties of PVC materials under the headings of mechanical properties, electrical properties and their resistance to ageing, moisture and various chemicals.

Mechanical properties

When considering those admixtures that affect the mechanical properties of PVC materials we shall confine our attention to that which is of the most importance — as its name implies — viz. the plasticizer.

The general rule is that a viscous plasticizer yields a harder PVC product than a more fluid one used in the same concentration. There are, however, quite a number of exceptions to this rule: sometimes the solvability, i.e. the property of saturating the secondary valencies of the molecules of the polyvinyl chloride, is of more importance than the viscosity of the plasticizer. There are, for instance, viscous, resinous, but highly solvating plasticizers which yield an astonishingly flexible PVC product.

The higher the content of a certain plasticizer, the less is the hardness, the modulus of elasticity, the tensile strength and the rigidity of the product; on the other hand the elongation at break increases. An example of this is given in *fig. 1*.

Elevation of temperature gives greater mobility to the molecules, the material becomes softer, more flexible and less strong, the deformation assuming more and more a plastic character. *Fig. 2* shows how the rigidity [3]) of PVC materials made with different plasticizers is affected by temperature. It is clear that in this respect a plasticizer like dioctyl sebacate is more favourable than, say, tricresyl phosphate.

[3]) Measured according to the specification D 747-43 T of the American Society for Testing Materials.

The fact that most of the (mechanical) properties of PVC are a function of temperature is due mainly to the viscosity of the mixture being temperature-dependent. Much has already been written about this temperature-dependency [4]). As a general rule it may be said that the more aromatic the character of the plasticizer, the greater is the dependency upon temperature. Examples of more aromatic plasticizers are tricresyl phosphate and dibenzyl phthalate. On the other hand, in the case of esters having long aliphatic chains, like dioctyl sebacinate, trioctyl phosphate and butyl acetyl ricinoleate, the viscosity changes but little with temperature.

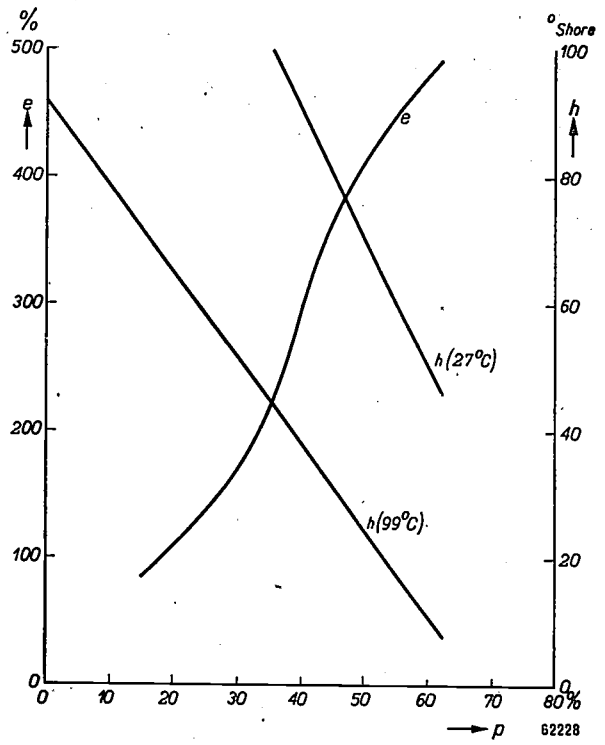


Fig. 1. The breaking strain e and the hardness h, expressed in °Shore, of a certain kind of "Podur", as a function of the plasticizer content p (p in wt % of the quantity of polyvinyl chloride).

Compared with the rubber that is used for insulating, PVC has greater abrasion and tear resistance. It is for this reason that in the U.S.A. for instance rubber has been entirely replaced by PVC for field telephone lines ("assault wire").

Duggan has compared the abrasion resistance of PVC with that of rubber by dragging segments covered with sandpaper over the material under a certain pressure. This test showed that the greater the content of plasticizer the greater is the wearing strength of the PVC product. From Duggan's figures it may be concluded that the kinds of "Podur" used as insulating material have 75 to 100% higher abrasion resistance than a standard rubber mixture answering the specifications D 395-40 of the American Society for Testing Materials.

As far as other properties are concerned good grades of PVC materials are on a par with or closely approximate the best kinds of rubber. From *table I* it appears that most properties are subject to considerable variations, but it should be borne in mind that the values given here cannot be combined arbitrarily; high elongation, for instance, cannot go hand in hand with great hardness.

Table I. Some mechanical properties [5]) of flexible PVC materials.

Tensile strength at normal temp. . .	$10 \cdot 10^6$-$25 \cdot 10^6$ N/m² [6]) (100-250 kg/cm²)
Rigidity at normal temp.	$3.5 \cdot 10^6$-$200 \cdot 10^6$ N/m² (35-2000 kg/cm²)
Hardness at normal temp.	60-200 °Shore
Elongation at break, at normal temp.	100-500 %
Permanent deformation under a compression load of $7 \cdot 10^5$ N/m² (7 kg/cm²), at 70 °C	10-15 %

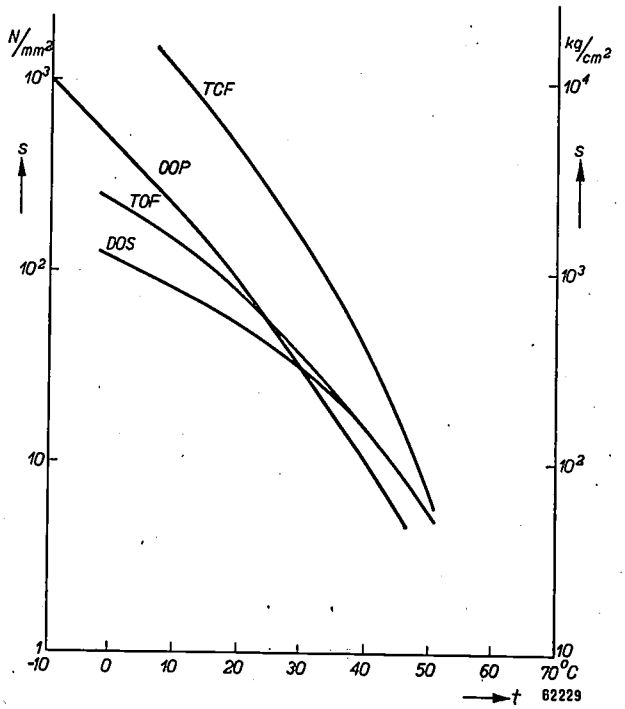


Fig. 2. The modulus of rigidity [3]) s of certain types of "Podur" with different plasticizers (content p = 40% in each case), as function of the temperature t. *TCF* = tricresyl phosphate, *DOP* = dioctyl phthalate, *TOF* = trioctyl phosphate, *DOS* = dioctyl sebacate. For the scale on the left see footnote [6]).

[4]) See, e.g., H. Jones, Trans. Inst. Rubber Ind. **21**, 298-322, 1946; E. M. Bried, H. F. Kidder, C. M. Murthy and W. A. Zisman, Ind. Eng. Chem. **39**, 484-497, 1947; R. T. Sanderson, Ind. Eng. Chem. **41**, 368-374, 1949 (No. 2).

[5]) Determined according to the specifications of the American Society for Testing Materials.

[6]) 1 N = 1 newton = 10^5 dynes ≈ 0.1 kg force.

Electrical properties

The most important electrical properties are conductivity, dielectric constant, loss angle and breakdown strength.

The conductance of PVC is to be ascribed to the presence of charge carriers, such as ions, which are capable of displacement over relatively large distances. Conductivity increases with the number and mobility of the ions, that is to say, according to the quantity of certain impurities contained in the polyvinyl chloride and in the plasticizer and according to the content of plasticizer and the temperature.

Investigations regarding potentials at the interfaces between inhomogeneous areas in PVC materials have given rise to the question whether a PVC material exposed to direct current does not in course of time change in its properties. Würstlin [7]) found, indeed, that under the influence of a direct voltage applied for a long time a highly diluted solution of polyvinyl chloride in different plasticizers splits up into two layers, one poor and the other rich in plasticizer. But, it should be added, Würstlin used for his tests a polyvinyl chloride that contained impurities, due to the method of manufacture, so that electrically it was less satisfactory.

The kinds of PVC behaving well in this respect, such as "Podur", have such little conductivity that even after ten years no change is to be expected. Following upon the American and German tests already mentioned, we have carried out durability tests with "Podur" wire. Neither under 600 V direct voltage nor under 220 V alternating voltage could any changes in the structure of the insulating material be found even after long exposure, a fact which would be expected considering the high insulation resistance of this material.

Not only the plasticizers but also the pigments influence the resistance of PVC. Some of them may, even in the small concentration of 0.5%, reduce insulation by a factor of 10 or even as much as 100.

Figs. 3 and *4*, relating to a particular case, give an idea of the manner in which the specific resistance of PVC depends upon the temperature and the content of plasticizer. In fig. 3 a curve is added for a good quality rubber, showing a temperature-dependency similar to that of PVC but lying higher. For almost all practical purposes however the insulation resistance of PVC is more than ample.

The dielectric properties of PVC are typically those of a polar plastic. Particularly the polyvinyl chloride molecules contain dipoles in large numbers, viz. bonds of carbon and chlorine atoms. When the PVC is exposed to an electric alternating field then,

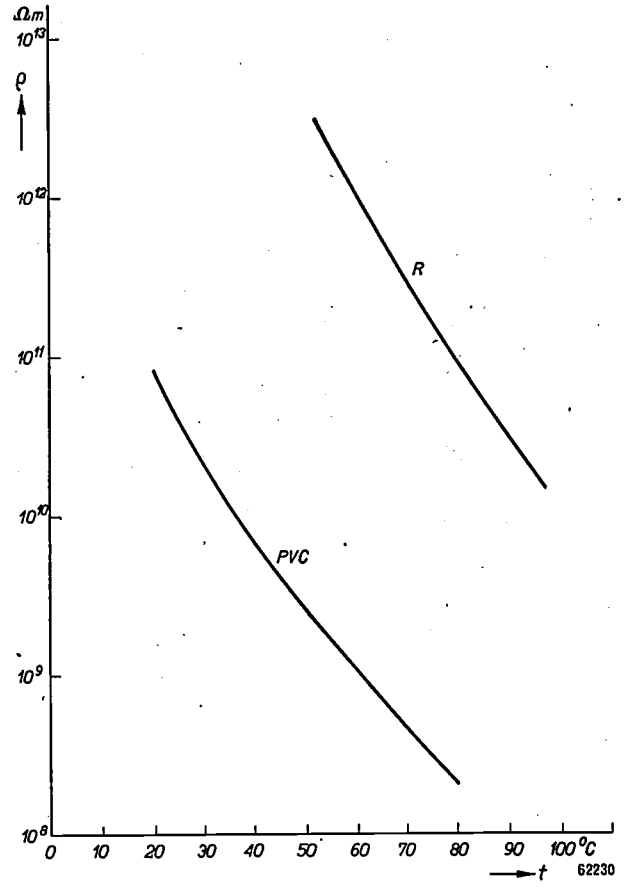


Fig. 3. Specific resistance ϱ of a flexible kind of PVC (*PVC*) and of rubber (*R*) of comparable flexibility, as function of the temperature t.

owing to their limited mobility, the dipoles lag behind the alternations in the direction of the field (after-effect [8])). Corresponding to this lag is a phase shift δ between the field strength and the dielectric displacement, which means that losses occur in the dielectric.

In the simple case where there is only one relaxation time τ to be reckoned with, it can be calculated that tan δ as a function of the frequency f of the alternating field has a maximum at $f = 1/(2\pi\tau)$. With most materials, however, there are several relaxation times, differing, moreover, in their dependency

[7]) F. Würstlin, Kunststofftechnik 11, 269-272, 1941.

[8]) See J. L. Snoek and F. K. du Pré, Several after-effect phenomena and related losses in alternating fields, Philips Techn. Rev. 8, 57-64, 1946, and M. Stel and E. C. Witsenburg, Heating by high-frequency fields, II. Capacitive heating, Philips Techn. Rev. 11, 232-240, 1950 (No. 8), in particular under "Nature of the dielectric losses".

upon temperature. Circumstances may be such that at one particular temperature, in a wide frequency band, tan δ changes but little, whereas at other temperatures it shows one or more maximum values. And this is in fact what is observed in the case of PVC materials (*fig. 5*).

The graphs for the dielectric constant show a less freakish picture, falling gradually with increasing frequency and rising gradually as the temperature rises (*fig. 6*).

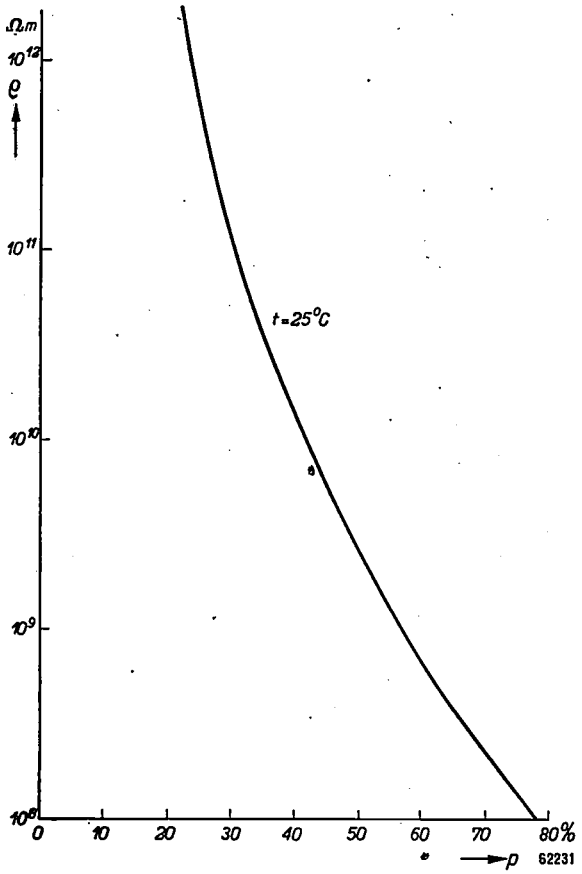


Fig. 4. Specific resistance ϱ of a kind of PVC as function of the plasticizer content p at 25 °C.

From figures 5 and 6 it follows that, generally speaking, PVC is not suitable as an insulating material in tuned circuits: tan δ (in the order of 0.1) and the relative dielectric constant (averaging about 5) are both too high and too temperature-dependent. Very little can be done about this by choosing a different plasticizer, because the dielectric losses are mainly due to the chlorine-carbon dipoles of the polyvinyl chloride itself. These dipoles are far more numerous than those of the plasticizers to be considered, so that it would serve little purpose to use a less polar plasticizer.

The dielectric strength (voltage gradient at which disruption takes place) has been plotted in *fig. 7* as a function of temperature. It is seen that it becomes less with rising temperature and that it is greater for direct voltage than for alternating voltage (even when reckoning with the peak value of the alternating voltage).

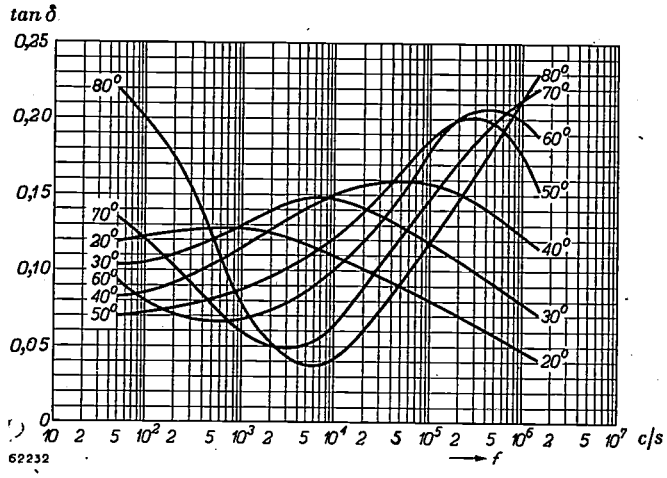


Fig. 5. Loss factor tan δ of "Podur" as function of the frequency f at different temperatures t (20 - 80 °C).

Resistance to ageing, moisture and chemicals

The polyvinyl chloride molecule stands up well against many anorganic acids, alkalies and salt solutions. The molecule is moreover saturated, that is to say it does not contain any double bonds (contrary to rubber for instance). The latter feature is favourable as regards durability but on the other hand it means that polyvinyl chloride cannot be vulcanized in the classical way [9]).

[9]) By classical vulcanizing is to be understood the conversion of double bonds into single ones, the free valencies being used for forming cross bridges to neighbouring molecules with the aid of sulphur. Vulcanizing reduces the thermoplasticity of the material.

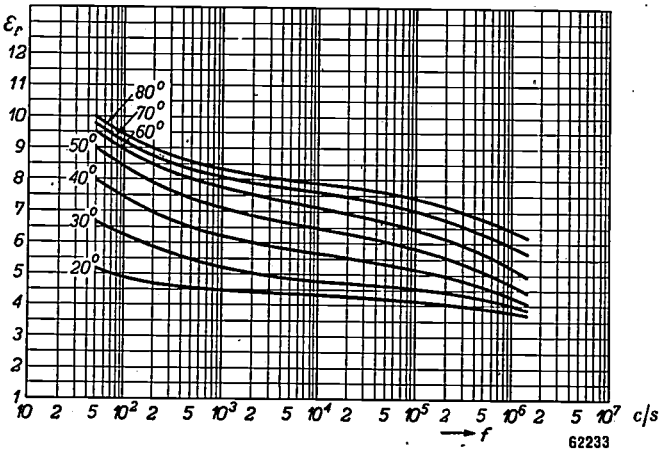


Fig. 6. Relative dielectric constant ε_r of "Podur" as function of the frequency f at different temperatures t (20 - 80 °C).

Greater danger originates from the admixtures. Provided these are well chosen, however, PVC is proof against oxygen and ozone in concentrations which cause rubber to age very quickly. Under the influence of ultra-violet rays there is some danger of decomposition, but, as already stated, this can

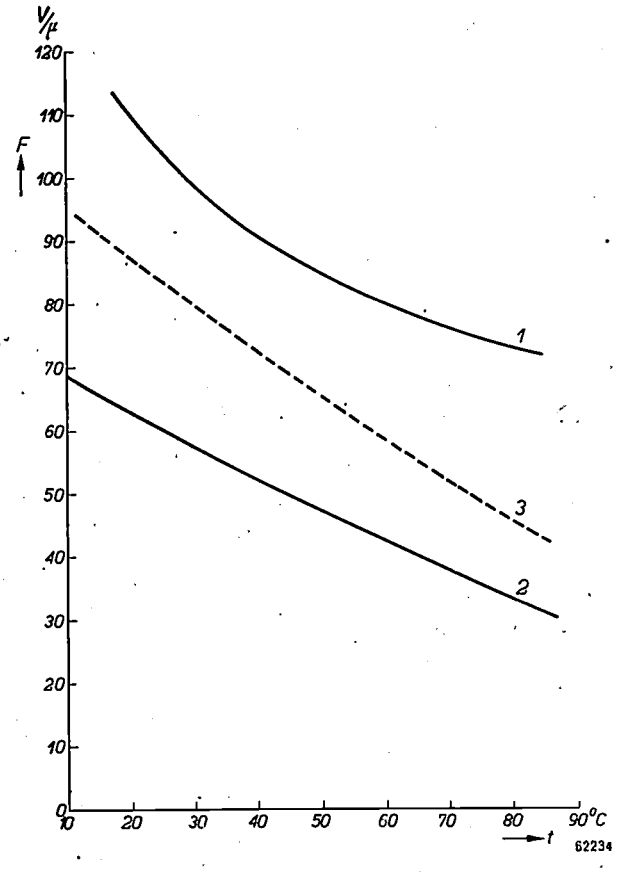


Fig. 7. Voltage gradient F at which breakdown takes place in "Podur", as function of the temperature t. Curve *1* is for a direct voltage, curves *2* and *3* for an alternating voltage of 50 c/s (*2* gives the R.M.S. value, *3* the peak value of F). At each test the gradient was raised to the breakdown value in 20 seconds.

be counteracted by adding certain stabilizers. The right choice of these substances depends largely on the purity of the plasticizer and the nature of the pigments.

In the introduction mention has been made of the non-inflammability of PVC, a most important advantage compared with rubber.

Most plasticizers are esters and therefore saponifiable (by saponification is understood here the reformation of the alcohol and the acid used to make up the ester). Saponification yields volatile products with unfavourable electrical properties. The rate at which it takes place differs widely as between one plasticizer and another. Tricresyl phosphate for instance saponifies very slowly, but PVC materials made with this plasticizer do not stand up so well against light as those made with phthalates.

Some organic substances cause more or less swelling in PVC materials. After evaporation of the swelling agent the material resumes its original shape and mechanical strength. Use is made of this phenomenon for covering objects with a protective layer: an extruded tube of PVC is caused to swell in acctone, which has a very strong swelling action, and then passed over the object. After the swelling agent has been evaporated the PVC tube shrinks tightly round the object and forms a strong covering layer of good appearance.

A good quality PVC material stands up excellently against moisture. Long immersion in water or burial in the ground has practically no effect upon the electrical and mechanical properties, as proved, for example, by the test mentioned in the introduction where a cable after having been left in the ground twelve years was still in very good condition.

This resistance to moisture does not mean that PVC is impermeable to water. For a cable consisting only of one or more cores and PVC insulation the penetration of moisture — which as a matter of fact takes place only to a very small extent — is quite harmless. It is a different matter, however, with a cable in which the cores are mutually insulated with a hygroscopic material like paper; the usual watertight lead sheathing may certainly not be replaced off-hand by PVC. Incidentally it is to be noted that there are synthetic products based on polyethylene or polyisobutylene which are less pervious to water than PVC, but these are beyond the scope of this article.

Mention should also be made of the fact that PVC and copper do not react upon each other, so that bare copper wire can quite safely be coated with PVC; such is not the case with rubber, since rubber and copper show a strong interaction, which is the reason why they have to be separated, for instance by using tinned copper wire. For the sake of easy soldering, however, tinned copper wire is sometimes used also with PVC insulation, as in the case of "Podur" hook-up wire.

It has already been remarked that when PVC is in contact with certain plastics, migration of the plasticizer is apt to take place. This may be harmful both for the PVC and for the contacting material (which of the two suffers most depends in the first place upon the relative quantities). It occurs particularly when PVC is in contact with polystyrene [10])

[10]) Polystyrene is better known under various trade names, such as "Distrene", "Lustrex", "Lustron", "Styron", "Trolitul", etc.

and with cellulose nitrate, especially with those kinds of the latter which have a small molecular weight, such as the cellulose nitrate used in the lacquer with which metal lamp fixtures, wooden radio cabinets and suchlike are sprayed. In unfavourable cases it is not long before so much of the plasticizer migrates as to cause the PVC to stick to the lacquer, much to the detriment of its mechanical strength ("lacquer lifting" or "surface marring").

The remedy against this migration is to use plasticizers with heavier and thus less readily displaceable molecules. It has been found, for instance, that polyesters are much less liable to migrate than non-polymerized plasticizers.

Table II gives comparative figures for the migration of different plasticizers into some plastics contacted with PVC. It is seen that an improvement with respect to one plastic does not necessarily mean an improvement for others: trioctyl phosphate, for instance, is very little absorbed by polystyrene but very much more so by cellulose nitrate.

Table II. Comparative figures for the migration of different plasticizers from PVC into various plastics in contact with it. The softeners are listed in the order of decreasing migration into polystyrene.

Plasticizer	Plastic in contact with PVC			
	poly-styrene	cellul. nitrate	cellul. acetate	etbyl cellulose
dibutyl sebacate	170	140	38	55
cresyldiphenylphosph.	168	253	16	41
dioctyl sebacate	163	113	0	21
benzylbutyl phthalate	159	158	18	19
tricresyl phosphate	105	154	3	0
dibutyl phthalate	91	162	74	50
dioctyl phthalate	78	88	70	7
trioctyl phosphate	43	270	18	19
polyester 3	22	76	0	0
polyester 1	14	51	5	0
polyester 2	0	121	1	0

Further it should be noted that there are also certain plastics — polythene is an example — which apparently do not absorb any plasticizer at all from the PVC, at least not in any chemically measurable quantity. And yet a gradual increase of the dielectric losses proves ummistakably that something does nevertheless happen. A typical example of this is found in the case of cables used in the beginning of radar technique. These cables had a polythene dielectric and a protective coating of "ordinary" PVC (with light plasticizers). It was noticed that the losses in these cables were getting greater and greater until ultimately they became prohibitive. This evil was overcome by changing over to cables coated with a PVC material in which exclusively heavy plasticizers were used.

What has been said above about the migration of the plasticizer into plastics applies in the main also for migration into oil. Oil does not react upon or dissolve the polyvinyl chloride itself, but in course of time it extracts the plasticizers. Again with oil, a great improvement is obtained when heavy plasticizers are used, such as polyesters, synthetic rubbers, etc.

Finally, a few words about the tropic-proof qualities of PVC materials. From their behaviour with respect to moisture, elevated temperature, and ultra-violet rays, it can be concluded that the better kinds of PVC may indeed be said to be tropic-proof. When the thermoplastic properties are not overlooked and no cotton or rayon is used for the insulation — these textiles deteriorate rapidly in the tropics — the use of PVC results in an enormous advance compared with rubber, as proved by the experience gained with "Podur" in Indonesia.

By way of illustrating the fact that the weak points of PVC are by far outweighed by its strong points, the case may be mentioned of an American factory where highly aggressive substances are used, such as caustic soda, hydrochloric acid, chlorine and sulphuric acid. Sometimes also very high or very low temperatures prevail in this factory, together with a high degree of humidity, factors which can certainly not be considered as favourable for the durability of an electrical installation. The cables in this factory are laid in channels in the concrete floors. They are mostly insulated with PVC, but, owing to this material being rather scarce at the time, some of the cables were insulated with rubber. Within a year the rubber cables started to cause trouble, so much so in fact that they had to be replaced. But the PVC cables have shown no defects whatever even after three years.

Coating wire with PVC

It has already been described how PVC is obtained in the form of chips or strands for feeding into the wire-coating machines. These machines, like those which produce the strands, are based on the principle of extrusion. A cross-sectional diagram of such a wire-coating machine is given in *fig. 8*. A worm rotating in a heated cylinder conveys the PVC from the feeder to the crosshead.

On its way the material is gradually heated, so that upon reaching the crosshead it is in a plastic state. Inside the crosshead it passes round a hollow pin called the mandrel or torpedo, through which the wire to be coated (bare or tinned copper wire) is drawn. The point of the mandrel reaches just into the die or outlet of the crosshead, where the coating takes place. By means of set screws the mandrel can be placed exactly concentric with

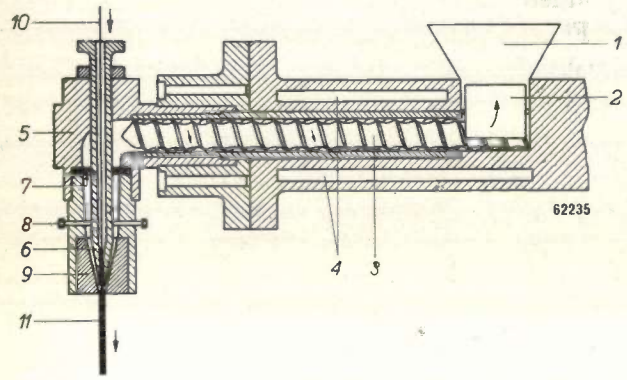


Fig. 8. Sectional diagram of a wire-coating machine. *1* filler, *2* drum carrying the chips of PVC to the conveyor worm *3*, *4* channels containing heating elements, *5* crosshead (with mandrel *6*, filter plate *7*, centering pins *8* and die *9*). *10* bare wire, *11* insulated conductor.

respect to the die, so as to ensure that the insulating layer is of the same thickness all round. *Fig. 9* is a photograph of one of these machines.

Careful attention has to be paid to the temperature variation of the PVC in this machine. If temperature is too low the surface of the insulated wire is rough and the material no longer passes the heat shock test described below. If the temperature is too high then again the wire is not of good appearance and, moreover, there is a risk of decomposition. Decomposition also arises in dead corners where the material is exposed to a rather high temperature for some length of time, so that it is of great importance that there should be no such dead corners where the PVC can stagnate.

The coated wire has to be tested for eccentricity. For those kinds of wire where a slight eccentricity does no harm it is specified that the breakdown strength must nowhere drop below a certain limit. In the testing machine the wire is passed at a high speed through a salt solution, with a high potential difference (say 6000 V) between the bath and the core. As soon as a breakdown takes place the testing machine is automatically stopped, this being done so quickly that the defective spot on the wire can easily be traced before having reached the take-up reel.

Wire intended for telephone cables needs a sharper test. The slightest eccentricity causes capacitive asymmetries in the cable, which result in cross talk. For the testing of these wires an apparatus has been devised through which the wire is passed as it comes straight from the coating machine and in which the slightest eccentricity is immediately detected, this then being corrected at once by means of the mandrel-adjusting screws, without having to stop the coating machine. This is a great advantage, because the manufacture can continue without interruption. When the machine had to be stopped, as was the case with the testing method formerly applied, there was a danger of decomposition of the PVC owing to its being left too long in the machine at a high temperature; once that process had begun it tended to spread through the mix, thus very soon making it necessary to give the whole of the machine a thorough clean-out.

Tests to which the coated wire is further subjected will be dealt with in the next section of this article.

When no bare wire is fed through the crosshead and the mandrel is advanced so as to protrude a little beyond the die then, as is quite obvious, the machine produces PVC tubing. Sleeving for mounting purposes is made in this way.

Types of "Podur" products

Without any attempt at giving such complete details as may be found in a catalogue, a description is given of the most important "Podur" products, with their fields of application and the main requirements that have to be satisfied in each particular case.

Connecting wire and flex

In its simplest form "Podur" connecting wire (including hook-up and applicance wire) consists of tinned copper wire with a single insulating layer of PVC. A coating 0.4 to 0.6 mm thick suffices in order to meet all the requirements specified by Philips Material Testing Department for wire carrying voltages up to 1000 V A.C. or 1500 V D.C. It would lead us too far afield to go into all these requirements, but two of them may well be mentioned here:

1. After 24 hours immersion in water at 20 °C
   a) the breakdown voltage, measured in mercury, must be higher than 7000 V A.C. (R.M.S. value) (actually it lies around 20,000 V), and
   b) the insulation resistance of 1 metre of wire

at 20 °C must exceed 100 megohms and 1 megohm at 70 °C.

2. After two hours in an oven at 140 °C the insulation must not show any cracks and the breakdown voltage at 20 °C must be higher than 7000 V A.C.

This second requirement might lead to the wrong conclusion that this wire can be used also at a temperature of 140 °C. Although the properties of PVC are unchanged after it has cooled down again to room temperature, a high working temperature is not permissible because then the PVC gets soft and it needs but little force to remove it from the wire. The heat-shock test specified is intended only as a test for the extrusion; if the extrusion is not carried out under exactly the correct conditions stresses arise in the PVC and when the heat-shock test is applied these stresses cause cracking.

The connecting wire is intended for fixed electrical connections inside electrical apparatus. If there is a possibility of the wire having to be mounted at a low ambient temperature it is not wise to use the hardest kind of "Podur", because in time the insulation is apt to break as a result of the repeated bending to and fro. The lowest temperature at which the wire can be used without this risk arising depends not only upon the kind of "Podur" but also upon the core diameter d. In order to investigate this question we constructed a simple device with which bending tests can be carried out in a reproducible manner and at different temperatures. These tests show that if it is required that the insulation should not break before the copper core itself, wire with $d \geqq 1.0$ mm and insulated with hard "Podur" may only be used for mounting at temperatures above 10 °C, and that wire with $d \leqq 1.8$ mm and insulated with moderately soft "Podur" can still be used below a temperature of 0 °C.

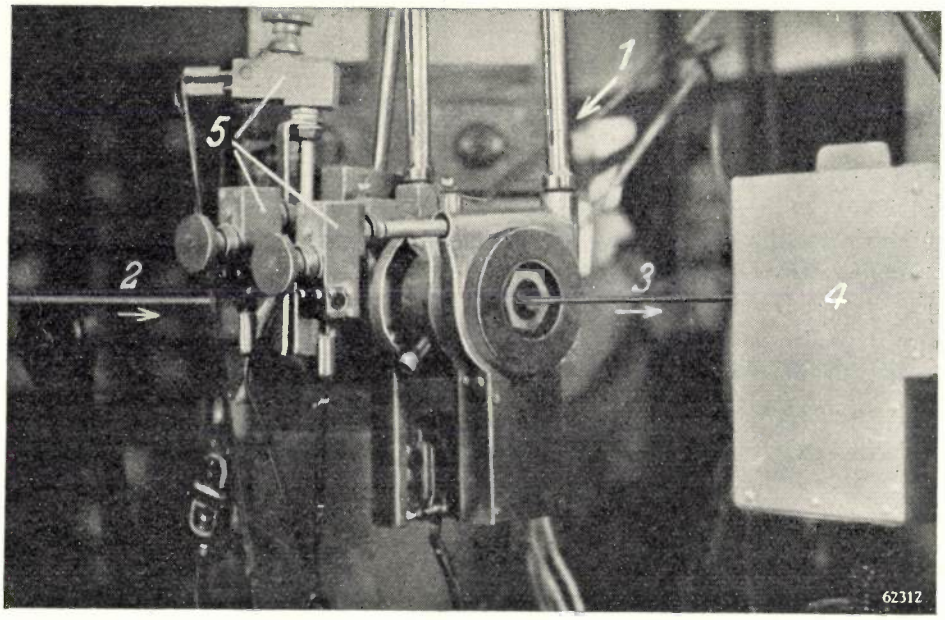


Fig. 9. Photograph of a wire-coating machine fed with a strand of PVC (*1*) instead of chips. *2* the bare wire, *3* the PVC-insulated wire, which is cooled in *4*. *5* are the temperature control instruments.

Where a flexible connection is required in an apparatus preference should be given to mounting flex with a stranded core consisting of a number of soft, tinned copper wires. One example of the use of such a flex which might be mentioned here is the illumination of Christmas trees, for which purpose the PVC-insulated flex is ideal, because it is non-inflammable, is more durable than rubber (so that it can be used for a great many years), is quite safe as regards dielectric strength (breakdown voltage >1000 V A.C.) and can be made in a suitable colour; usually it is green, but sometimes preference is given to transparent PVC flexes.

In addition to the single-core flex, flexes are also made with two insulated cores twisted together. Both the single- and the twin-core flex can be screened with a copper wire braiding, with or without

an outer sheathing of flexible "Podur". These constructions, however, are not suitable for high frequencies.

Connecting wire is also made with "Podur" insulation thicker than 0.5 mm, say with a thickness of 1.6 mm, which is suitable for working voltages up to 6000 V A.C. or 10,000 V D.C.

With all these kinds of mounting wire and flex, if there is any chance of a high temperature being reached, care must be taken to avoid the possibility of any force being brought to bear upon the wire as this would be likely to push the insulation away.

Insulating sleeving

"Podur" sleeving answers the American specifications (American Society for Testing Materials, D 876 and D 922) as regards insulation resistance, dielectric strength, breaking stress, tensile strength, resistance against oil and ageing, while it also just answers the "penetration" test, which requires that the thermoplasticity must not be so great that above a certain temperature and at a given pressure the core would be exposed.

In cases where it is difficult to avoid conditions whereby the insulating material might be pushed away it is often possible to make allowance for this by using a thicker coating or a harder kind of PVC.

"*Parallel connecting flex*"

For the connecting leads of radio sets, reading lamps, etc. flexes are made with two parallel cores (not intertwisted). This parallel-core connecting flex is available in various colours, offering a choice of shades matching any surroundings.

A special type of this kind of flex has been made for the "Philishave" dry-shaver (*fig. 10*), which has to be very flexible to stand up against the daily rolling up and unrolling, whilst moreover, since it is frequently in contact with the hands, the insulation has to answer very high requirements. It is required, for instance, that after 48 hours in an atmosphere of 94% relative humidity the insulation resistance of 1 metre of each core at 20 °C has to be greater than 100 megohms, and after 48 hours immersion in water the breakdown voltage between cores has to be higher than 5000 V A.C. (the permissible working voltage is 500 V A.C. or 750 V D.C.). The measured insulation resistance of 1 metre is about 2000 megohms. The condition specified in Switzerland, that the leakage current of 2 metres of cable in water of 50 °C to which an A.C. voltage of 500 V is applied must be less than 0.5 mA, is amply satisfied. (This requirement was based on the consideration that a current of 5 mA passing through the human body may cause such cramp as to make it impossible to release the grip on the cable, and that the first symptoms of shock occur at a current of 1.7 mA peak value.)

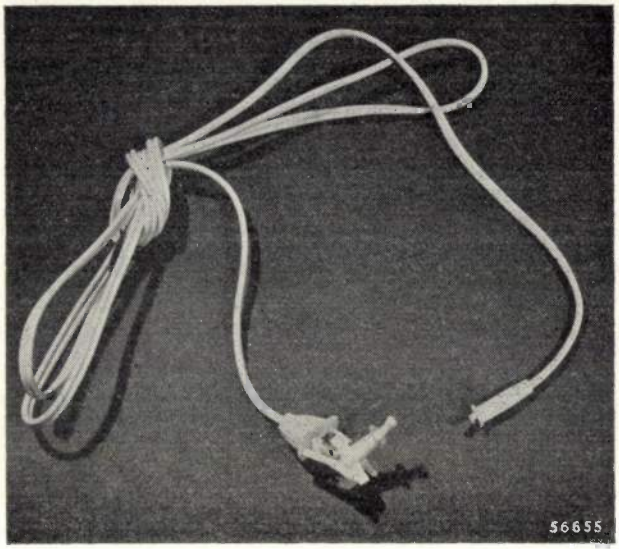


Fig. 10. "Podur"-insulated flex for the "Philishave" dry-shaver.

Flex (or cable) with two or more intertwisted cores

Flexes or cables with 2, 3, 4, 7, 12 or 19 intertwisted cores, with core diameters varying between 0.75 and 6 mm^2, are manufactured for all sorts of purposes.

Not only the cable as a whole but also each individual core has to satisfy certain demands as regards insulation resistance, breakdown voltage and loss angle. With separately tested and approved cores it has been possible to build up highly dependable cables, only the cores themselves having to be rejected where necessary.

If necessary these cables can be supplied with a shielding and a PVC outer sheath.

High-tension cable for television tubes

A thin cable was required for a television receiver [11]), for connecting the cathode ray tube to the supply unit delivering an anode voltage of 25 kV. It was specified that this cable had to withstand a D.C. voltage of 40 kV for two minutes at 65 °C and that a length of 1 metre should show an insulation resistance of more than 1000 megohms. The insulating material has to be resistant to oil (the parts of the supply unit are contained in a can filled with oil) and it has to make a seal tight

[11]) G. J. Siezen and F. Kerkhof, Projection-television receiver, III. The 25 kV anode voltage supply unit, Philips Techn. Rev. **10**, 125-134, 1948.

enough to prevent oil leaking out between the core and the insulation. These requirements have been met by making a cable (*fig. 11*) with a thick coating of "Podur" of a kind specially prepared for the purpose (outer diameter 4.4 to 4.7 mm). The ingredients for the mixture and the manner of processing have to be carefully tested.

Telephone cable

Telephone cables (not intended to be laid underground) are being made with 3 to 37 groups of cores, each group consisting of two, three or four conductors. Each core is insulated with "Podur" in different colours (red, blue, yellow, green, black and white) according to a specified colour scheme. The intertwisted cores are sheathed with grey "Podur".

Since it would be impracticable to give such large numbers of cores each a different colour, two or more colours are used per core. The additional colours are printed on in the form of helical lines round the core, for which a special machine has been designed.

Mention has already been made of the precision with which the core is centered in the insulation in the manufacture of the coated wires. The slightest eccentricity gives rise to cross talk. The finished cable is likewise tested for such tendency (*fig. 12*).

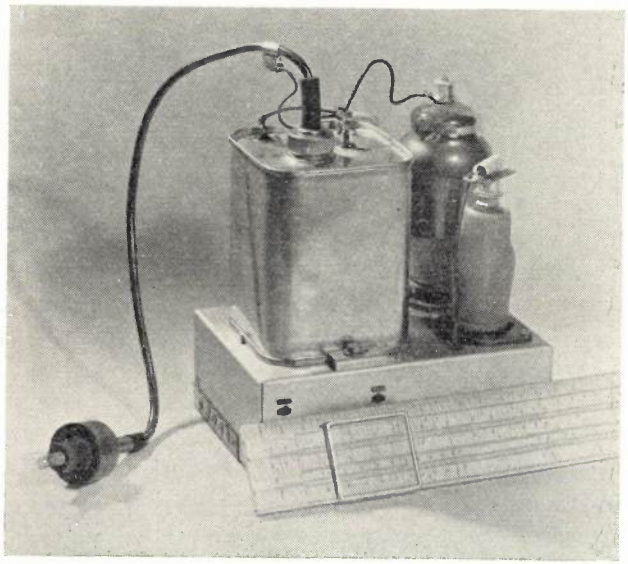


Fig. 11. Anode supply unit for supplying 25 kV anode voltage to the cathode ray tube of a television receiver. On the left the "Podur"-insulated cable connecting up to the cathode ray tube.

As regards the other inspection tests, briefly these amount to an insulation resistance for 1 metre of cable exceeding 10^4 megohms and a capacitance less than 100 pF/m at 25 °C between the two wires of each pair.

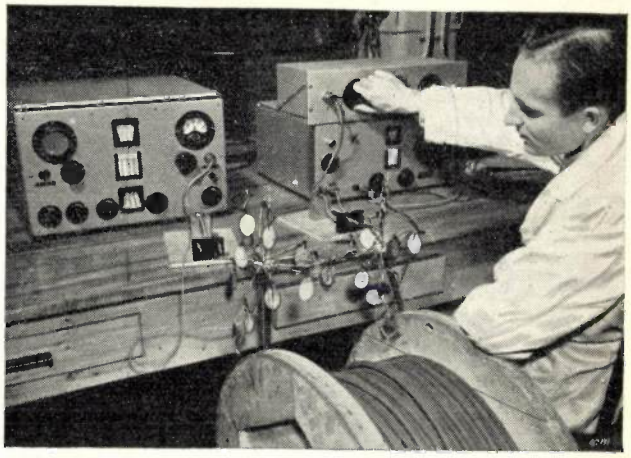


Fig. 12. Measuring the cross-talk damping between the pairs of conductors in a telephone cable. A certain voltage supplied by the generator (left) is applied to one pair of conductors. The resulting voltage on other pairs is measured (right) and must not exceed a certain value. Measurements are taken at different frequencies (100 to 500 kc/s).

Microphone flex for telephones

The flex used for connecting the microphone to the other parts of a telephone has to be particularly flexible and must withstand bending a large number of times.

With a view to obtaining the desired flexibility, so-called tinsel is used (a thin strip of copper wound round a thread of artificial silk), and for the insulation of the wires an exceptionally flexible kind of PVC (red and blue, yellow and green for the different cores, which are four, three or two in number). Notwithstanding its great flexibility, the insulation resistance of this cable, per metre, still exceeds 100 megohms as tested at 20 °C and after 6 hours at 90% humidity.

The cores are intertwisted and braided with cord. The breakdown voltage between the cores has to be higher than 1000 V, and after the flex has been bent to and fro 6000 times there must not be any "crackling".

Microphone flex for amplifiers

"Podur" microphone flex has two or four conductors, which, since resistance is of no importance, may be thin (0.3 mm in diameter; they are made of cadmium-copper for strength). The cores are separately insulated with PVC, intertwisted, screened and once more insulated with a sheathing of PVC. This cable is tested for insulation resistance, screening of the cores and microphony (the occurrence

of potential differences in the cable due to bending or twisting).

Flex for hearing aids

Modern hearing aids consist of a small box containing a microphone, amplifier, and batteries and a miniature telephone worn in the ear. They are usually required to be as inconspicuous as possible.

The flex connecting the telephone to the amplifier has been made inconspicuous by using a very thin coating of a colourless, transparent PVC, with an outer diameter of no more than 1.6 to 1.7 mm. It has two very flexible conductors ("tinsel"). Owing to the small dimensions — calling for the utmost precision in manufacture — and the high flexibility of the PVC, a very flexible lead has been obtained which causes no inconvenience whatever to the wearer of the hearing aid.

This summary must suffice for the present. It is hoped that it has sufficiently proved that PVC insulating material has already acquired a position in many fields.

Summary. Following upon a discussion of the components used for preparing wire-insulating materials based on the synthetic resin polyvinyl chloride, some of the properties of these materials are dealt with under the headings of mechanical properties, electrical properties and resistance to moisture, chemicals, etc. Some of these properties are superior to those of rubber (i.e. non-inflammability, abrasion resistance, resistance to chemicals and ultra-violet rays), while others closely approximate them (insulation resistance). Attention is drawn to certain points that have to be watched when using insulating materials, such as thermo-plasticity. An account is given of the extrusion process, by means of which wire is coated with polyvinyl chloride insulation. In conclusion a brief description is given of the principal types of wires with polyvinyl chloride insulation ("Podur") marketed by Pope's Wire and Lamp Works at Venlo (Holland): connecting wire and flex, various special flexes, high-tension cable, telephone table, microphone flex, etc.

INACTIVATION OF BACTERIA BY ULTRA-VIOLET RADIATION

by J. VOOGD and J. DAAMS.

621.327.3:613.165.6:
614.7-084.48:664

Radiation from the sun has different effects upon living organisms. It has been found, for instance, that ultra-violet rays of a certain wavelength present in the radiation from the sun stop the multiplication of bacteria and other micro-organisms. Good use can be made of this, without having to depend upon sunlight, by employing a low-pressure, mercury-vapour lamp emitting rays which have a specifically disinfecting power exceeding by far that of sunlight. This lamp presents good possibilities for the solution of some problems of hygiene and for the sterilisation in processing foodstuffs and medicaments.

The fact that millions of pounds worth of foodstuffs are spoiled every year through bacterial action shows how important it is that the growth of these micro-organisms should be checked also in this field.

The influence of sunlight upon bacteria

In the year 1877 the Englishmen Downes and Blount made a discovery of great value for the existence of the human race. They found that micro-organisms cease to multiply when exposed to radiation from the sun for a sufficient length of time. The effect of such exposure can be demonstrated by a simple experiment, the result of which is represented in *fig. 1*, where we see a so-called Petri dish, such as is commonly used by bacteriologists. First a culture medium particularly suitable for making the growth of Bacterium coli visible was placed in this dish. Where the bacteria colonise this culture medium assumes a deep red colour owing to the formation of fuchsin, but where there is no growth of the bacteria it remains clear. A suspension of coli bacteria was seeded all over the surface of the medium in this circular dish. The left half was then exposed for one hour to direct sunlight falling upon it at an angle from the left, whilst the right half was covered with a screen intercepting the ultra-violet rays. After 24 hours' heating at 37 °C the dish showed the picture of fig. 1. The dark part of the medium in the right half indicates mass growth and colonisation of the bacteria. The light half on the left shows that there has been no growth of the bacteria there. The sharp demarcation between the two parts clearly demonstrates the effect of sunlight.

It is to be noticed that the light part does not extend right up to the edge of the dish. On the extreme left a number of strongly developed colonies are clearly perceptible, this being due to the shadow cast by the glass rim of the dish. This shadow effect is interesting, in that it indicates the answer to the question as to what part of the spectrum is most responsible for the inactivating action of the radiation from the sun. The fact is that ordinary glass fully absorbs all rays of a wavelength smaller than 3200 Å, so that — excluding the possibility of specific biological actions of the infra-red rays with wavelengths greater than 30,000 Å, which are likewise absorbed by the glass — the shadow effect of ordinary glass points to the conclusion that the inactivating action is to be ascribed to the small quantity of ultra-violet rays of wavelengths shorter than 3200 Å present in sunlight.

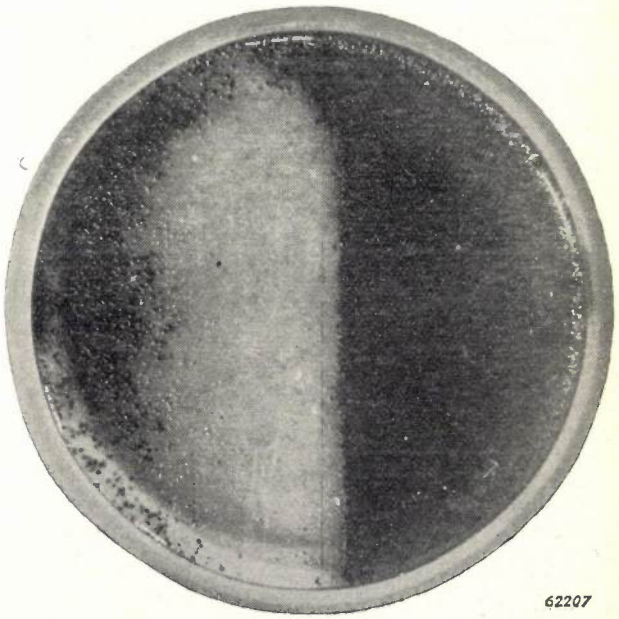


Fig. 1. Development of Bacterium coli. The left half of the dish has been irradiated with sunlight falling upon it at an angle from the left. Owing to the formation of fuchsin the culture medium where colonies have developed is of a darker colour.

The first question to be put is whether these rays act upon the bacteria or upon the culture medium, or upon both. The answer to this is given by an experiment the result of which is to be seen in *fig. 2*. A circular part of a culture medium sown

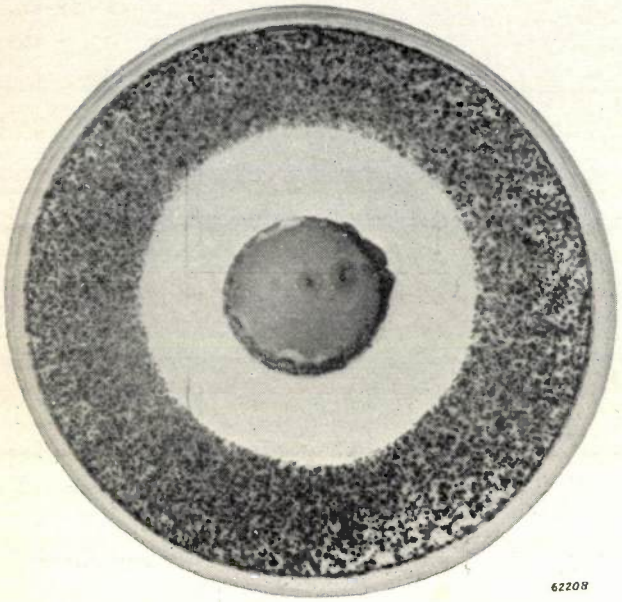


Fig. 2. Development of colonies of Bacterium coli on non-irradiated (outer ring) and irradiated (inner circle) culture medium, showing that the radiation has no effect upon the culture medium.

with bacteria was intensely irradiated, after which a piece in the middle of that part was re-sown. After heating it appeared that the non-irradiated bacteria in the middle of the irradiated circle are still able to develop on the previously irradiated culture medium. It is, therefore, to be concluded that the ultra-violet rays in sunlight have the power to stop multiplication of the coli bacteria. Briefly, though not absolutely correct, one speaks of the bactericidal action of sunlight. Analogous phenomena are observed when moulds, yeasts and viruses are irradiated, and consequently one speaks in more general terms of the germicidal action of ultra-violet rays.

Spectrum of the bactericidal effect

This discovery of the bactericidal effect of ultra-violet rays very soon led to an investigation into the question of its dependency upon the wavelength of the irradiation, a qualitative answer to which is given in *fig. 3*, which was produced in the following way.

A spectrum of the radiation from an electric discharge in mercury vapour under high pressure was projected, with the aid of a large quartz spectograph with wide aperture, upon the surface of a culture medium sown with a suspension of bacteria. For the radiation a "Biosol" lamp, type B, was used. After incubation the culture medium showed the picture given in fig. 3. Where the projected wide spectral lines are to be seen there was no growth of the bacteria, due to the action of the ultra-violet rays, such in contrast with the immediate surroundings where no or very little ultra-violet radiation reached the culture medium. Below this microbiologically recorded spectrum a photographically recorded spectrum of the lamp used has been reproduced, and below that the spectrum of the germicidal lamp which will presently be discussed.

In this experiment the strong spectral line of 3130 Å shows no bactericidal action, whereas such is the case with the spectral lines of smaller wavelengths.

Of course this experiment gives only a qualitative picture, from which no quantitative conclusions can be drawn regarding the dependency of the bactericidal action upon the wavelength of the radiation. To this end several investigators have taken careful measurements. *Fig. 4* gives curves plotted from information contained in literature on the subject.

Curve (*a*) shows a maximum at 2650 Å and applies with fairly great accuracy both for bacteria and for moulds. Attempts have been made to relate this spectral curve to an absorption process, and absorption curves have been sought which correspond to the bactericidal curve. It is remarkable that the spectral absorption curves of proteins, with a maximum at about 2800 Å, show no correspondence. Better agreement is found with the

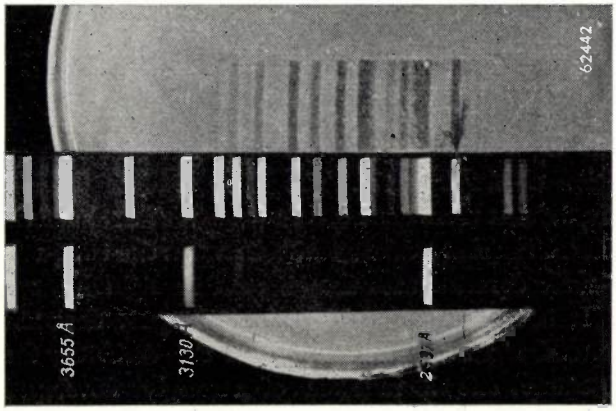


Fig. 3. Biologically recorded spectrum of the bactericidal effect of a "Biosol" lamp type B (*top*). Owing to the recording technique the lines where no bacteria have developed are darker. For comparison, the photographically recorded spectra of a "Biosol" lamp type B (*middle*) and a germicidal lamp, TUV 25 W (*bottom*).

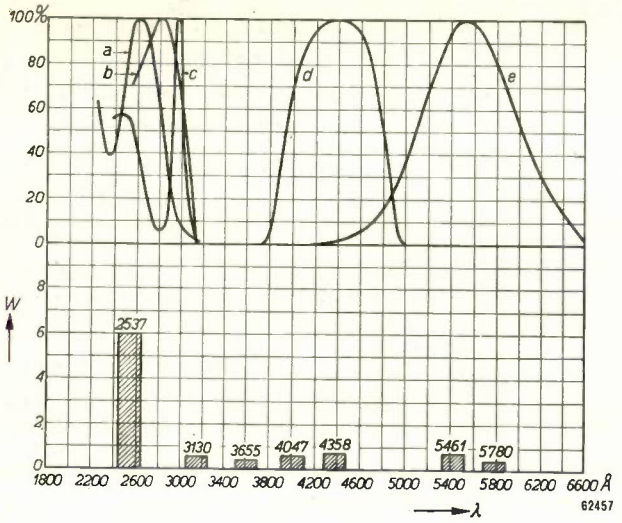


Fig. 4. *Top*: various effects of radiation as function of the wavelength.
a) Bactericidal action (B. coli).
b) Killing effect upon a protozoan (Paramecium multimicronucleata).
c) Erythemal effect.
d) Photographic effect (non-colour-sensitive film).
e) Normal eye-sensitivity.
Bottom: radiation from the TUV 25 W lamp.

spectral absorption of nucleoproteids (substances related to proteins from which, i.a., chromosomes are built up). This is demonstrated in *fig. 5*, where the spectral absorption curve of thymonucleic acid is given as an example of a nucleoproteid.

As regards the mechanism of the inactivation it can be proved that the metabolism in inactivated bacteria partly continues, but without leading to the normal division of cells.

In view of these phenomena the bactericidal effect of ultra-violet radiation is interpreted as an absorption process in those nucleoproteids performing a fundamental function in the multiplication of bacteria. Apparently this absorption process in itself is not deadly, the result of the irradiation being a normal dying off of the bacteria before multiplication has taken place (bacteriostasis).

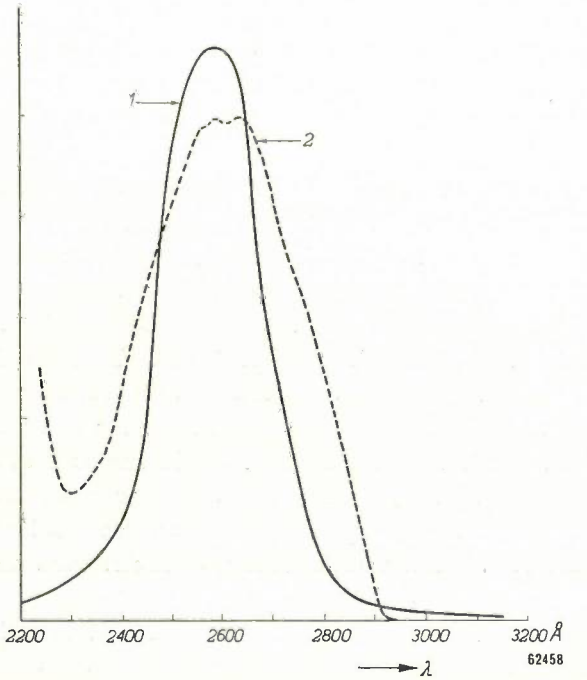


Fig. 5. Bactericidal action as function of the wavelength (*1*) and absorption curve of thymonucleinic acid (*2*).

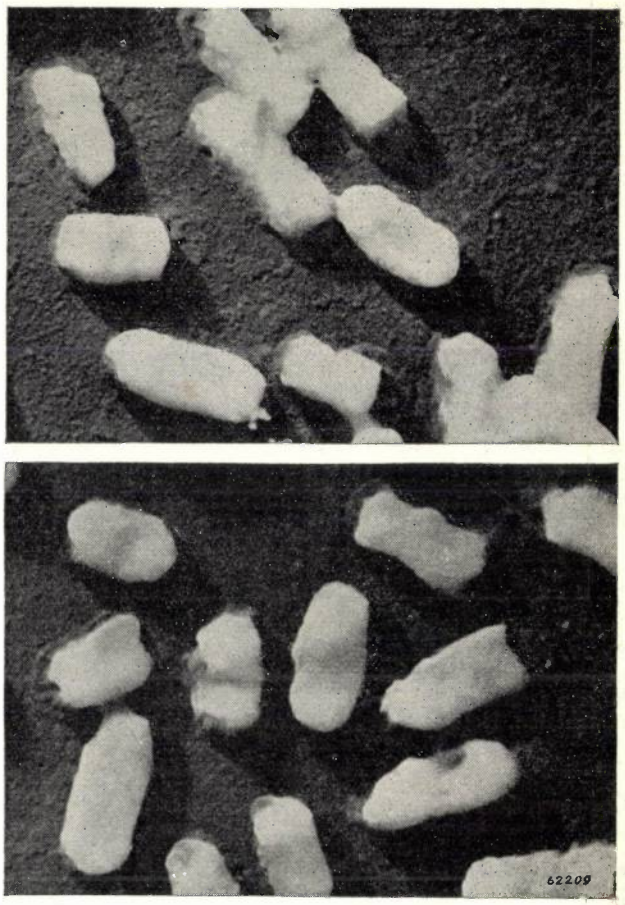


Fig. 6. Photographs taken with an electron microscope showing coli bacteria (enlarged 18,000 times). *Top*: not irradiated, *bottom*: irradiated with ultra-violet. Outwardly there is no difference to be seen.

From photographs obtained with the electron microscope it appears that when irradiating with a dose that undoubtedly stops multiplication of cells there is no outwardly perceptible damage to the bacteria, such as is the case in the destruction of protozoa (e.g. Paramecia) by ultra-violet irradiation. *Figures 6* and *7* give a good picture of this [1]).

[1]) It has also been found possible to restore inactivated bacteria to their normal state of multiplying — be it only in part — by powerful irradiation with visible light a short time after inactivation with ultra-violet radiation.

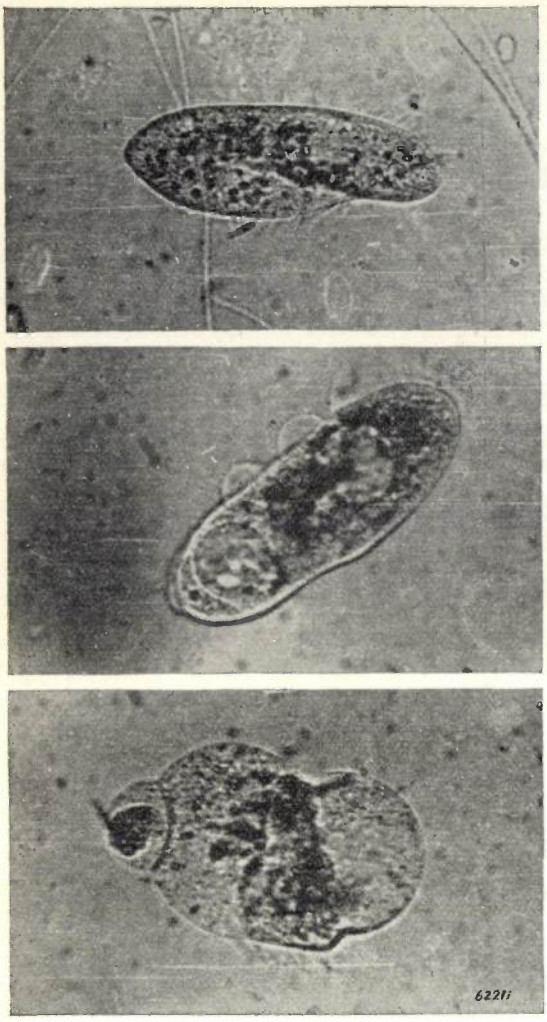


Fig. 7. Photographs taken with a normal microscope showing paramecia (enlargement 250 times). *Top*: not irradiated, *bottom*: irradiated with ultra-violet, in different stages of destruction. Note the blisters formed.

Construction of germicidal discharge lamps

With the modern technique in the manufacture of electrical discharge lamps it is possible to make a special lamp for the action described above, so that one is no longer dependent upon the sun for it.

As we have seen, radiation of a wavelength between 2500 and 2600 Å has a strong germicidal action. A powerful source of radiation within this wavelength range can be obtained by employing the electrical discharge in mercury vapour of low pressure, whereby mainly two resonance lines of the mercury atom are radiated, namely a weak one at 1850 Å and a strong one at 2537 Å.

A germicidal lamp can be dimensioned exactly like the tubular fluorescent lamp for lighting, in which ultra-violet radiation is used for producing visible fluorescent radiation in a powder with which the glass tube is lined.

For the germicidal lamp it is the ultra-violet radiation of 2537 Å that we want to use, that of 1850 Å being undesired on account of its strong ozonising action. Therefore in the manufacture of germicidal lamps the fluorescent powder has to be omitted and the tube has to be made of a kind of glass that passes the mercury line of 2537 Å but absorbs the lines of smaller wavelengths, and especially the resonance line of 1850 Å.

In this way a lamp has been constructed (the TUV 25 W) of which the spectral distribution of the radiation intensity is represented in fig. 4. Since this lamp, as regards the electrotechnical part, is exactly similar to the TL 25 W lamp, the same starter and ballast can be used for both.

The energy of the radiation in the wavelength of 2537 Å is 6 W, but this high output is only obtained when the vapour pressure of the mercury in the tube is of the right value, so that it is greatly dependent upon temperature. In view of the radiation of heat the tube has been so dimensioned that under normal conditions the wall temperature is about 40 °C, which is just right for the most favourable vapour pressure. When under different circumstances, e.g. when the lamp is mounted in a refrigerator the wall of the tube assumes a different temperature, the output is considerably reduced: in the case of deviations of 20 °C the output is about 25% lower.

Radiation doses required

For combating germs with the aid of germicidal lamps the first question to be decided is what power has to be installed.

Although there are a number of complicating factors to be reckoned with it may be taken as a basis that the degree of inactivation depends upon the product of the intensity and the duration of the irradiation.

The various species of bacteria, however, are not all equally sensitive to ultra-violet radiation, as may be seen from *table I*.

A very resistant bacterium for instance is Sarcina lutea, which is about seven times more resistant than Bacterium coli. Moulds are still less sensitive. *Fig. 8* shows how, in the case of a mixture of bacteria and moulds forming, when not irradiated, visible colonies on the culture medium, after irradiation with a suitably chosen dose the bacteria are rendered inactive while the moulds continue to multiply. It is to be taken that in practice for moulds a dose is required 20 to 100 times stronger than that needed for Bacterium coli.

Table I. Doses in mWsec/cm² required for inactivation of 90% of the following organisms (from various publications).

Organism	Dose	Organism	Dose
Bac. anthracis	4.52	Serratia marcescens	2.4
B. megatherium (veg)	1.13	ditto	2.2
B. megatherium (spores)	2.73	ditto	8.3
B. paratyphosus	3.20	Shigella paradysenteriae	1.68
B. subtilis	7.1	Spirillum rubrum	4.4
ditto	6.0	Staphylococcus albus	1.84
B. subtilis (spores)	12.0	ditto	3.30
Corynebacterium diphteriae	3.37	Staphylococcus aureus	2.18
Escherichia coli	3.0	ditto	2.60
Proteus vulgaris	2.64	ditto	4.95
Sarcina lutea	19.7	Streptococcus viridans	2.0

From table I it is seen that the values given by various investigators are widely divergent. This is due to the difficulties encountered in trying to obtain quantitative data on this subject. Even

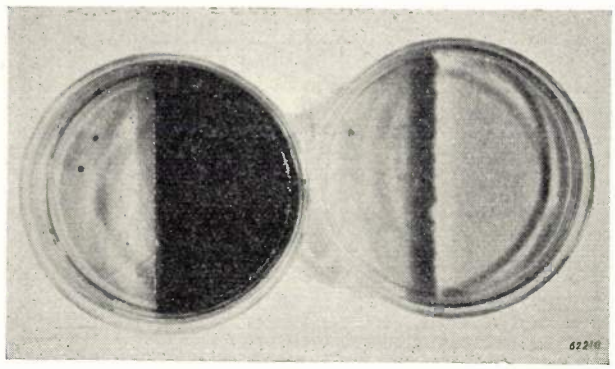


Fig. 8. Cultures on the left half irradiated with a TUV lamp of 25 W at 25 cm distance. On the left coli bacteria irradiated for 2 minutes, on the right a mould (Penicillium spores) irradiated for 15 minutes.

when attention is confined to one strain of a certain species of bacteria its sensitivity is found to vary according to the life phase in which the bacteria happen to be at the time of irradiation and also according to the nature of their surroundings.

Life phases of bacteria

As is the case with all living organisms, the vital process of bacteria is closely inter-related with the surroundings. This is clearly seen when a small quantity of an old suspension of bacteria is placed in a liquid culture medium, such as sterile broth. In such an old suspension the process of cell division is highly stagnated, due to the influence of the accumulation of metabolic products which inhibit growth of the bacteria in those surroundings. When these bacteria are placed in a fresh medium it takes some time for them to adapt themselves. In this adaptation phase the metabolic processes increase, the volume of the bacteria becomes greater and, as appears from excellent investigations of the Caspersson school [2]), also the content of nucleoproteids increases. As a result of these physiological activities the process of cell division starts anew. The culture then reaches what is called the logarithmic phase, where the number of bacteria increases almost exponentially with time; this phase brings about its own end by the inhibiting action of accumulated metabolic products in the suspension. In the bacteria changes take place which cause the process of cell-dividing to come to a standstill and consequently the natural dying off becomes a factor of importance. Via a stationary phase the culture then reaches the dying-off phase, where the number of bacteria decreases and ultimately the culture dies out.

The various conditions in which the bacteria are found in the ageing of the culture described above are found again in the process of inactivation by ultra-violet irradiation. This can be explained with the aid of *fig. 9*, relating to the following experiment.

A small quantity of a suspension of Bacterium coli in the dying-off phase was placed in sterile broth. The curve on the extreme left of fig. 9 represents the inactivation of this culture in the adaptation phase as a function of the dose of irradiation of 2537 Å. This and the other inactivation curves in the diagram were determined by sowing equal, small, quantities of the well mixed and shaken culture on culture media and then exposing them to irradiation for different lengths of time. After the necessary incubation the percentage was counted of the bacteria not yet rendered inactive. This percentage has been plotted in the graph on a logarithmic scale against the dose of irradiation.

From left to right, the curves each relate to a more advanced stage in the ageing of the culture sown in the medium. It has to be taken that this suspension reached the logarithmic phase in two hours.

It appears that in the state of the strongly aged culture (dying-off phase) the bacteria are less sensitive to ultra-violet radiation than in the logarithmic phase.

For practical purposes the curves given in fig. 9 show that it is wise to be careful when quoting figures for doses and that it is well to bear in mind that bacteria, and micro-organisms in general, may be in a more or less resistant state.

In addition to the differences in sensitivity something can also be deduced from these curves about the mechanism of the inactivation.

In cases where bacteria were inactivated by electron rays and alpha rays it appeared that one ionization in a specific part of the bacterium suffices. Supposing that, in analogy therewith, absorption of one quantum is needed to bring about inactivation by ultra-violet irradiation, then the number of bacteria inactivated per unit of time is proportional to the radiation intensity I and to the number of active bacteria N, thus

$$dN = -C N I \, dt.$$

By integration we get the relation:

$$\log_e \frac{N}{N_0} = -C I t,$$

and the graphical representation gives a straight line, as in fig. 9 *IV*.

[2]) B. Malmgren and C. G. Heden, Studies of the nucleotide metabolism of bacteria, Acta Path. Micr. Scand. 24, 417, 1947.

It may be taken that the absorption must take place in that nucleoproteid system that performs an essential function in cell division. Owing to the absorption a mutation takes place in the system, viz. a lethal one, whereby the capacity of multiplication is stopped. If a biochemical foundation could be given for this conception it would be of the greatest importance for our knowledge of bacteria.

The non-linear trend of the curves for the other phases is an indication that in those phases the ultra-violet inactivation is of a more complex nature.

required to reduce the number of bacteria to 1 - 0.1%, in a suspension atomized into the air the same effect was reached with a dose of 1 mWsec/cm².

This phenomenon, so important for practical purposes, is usually ascribed to a change of the micro-organism after atomization in air. We are still in the dark about the physiological aspect of this change. There is, however, a purely physical

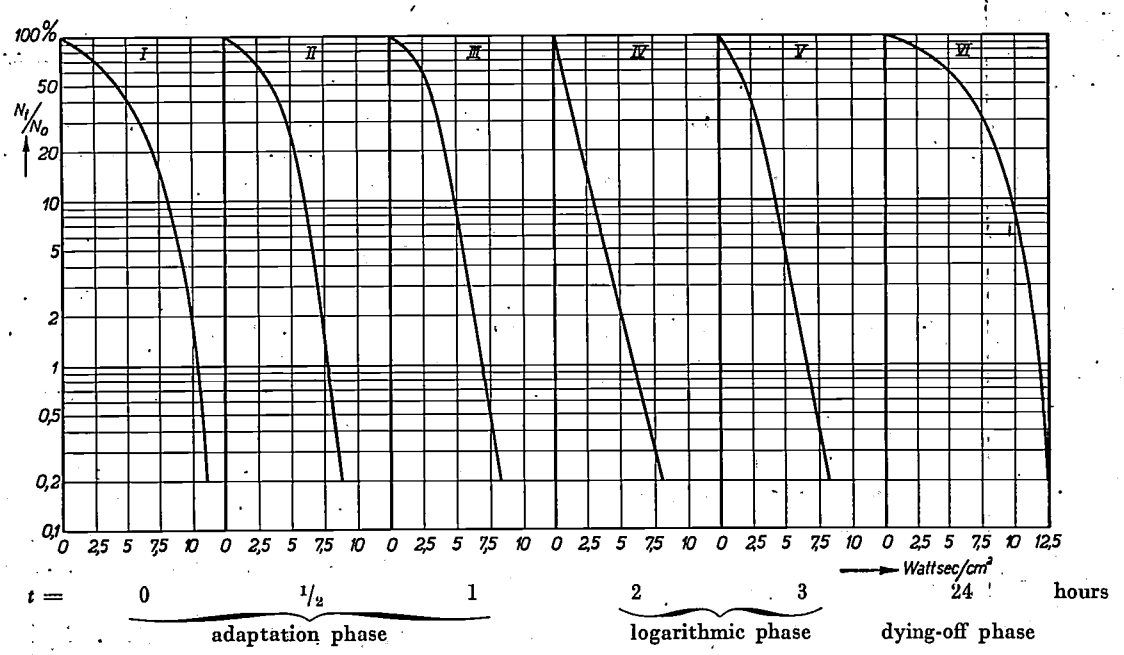


Fig. 9. Inactivation of coli bacteria in different life phases (*I* to *VI*) by irradiation with 2537 Å, as function of the dose (in wattsec/cm²). Along the ordinate the ratio of the number of bacteria not yet inactivated (N_t) to the original number (N_0). The dose as administered t hours after sowing the bacteria in the culture medium.

Influence of the nature of the medium

One of the most common applications of the germicidal lamps is for the sterilization of air, either stationary or in motion. Investigations have shown that the doses required to reach a certain degree of inactivation of germs in the air are very much smaller than those that follow from measurements taken in suspensions and from cultures on agar plates.

Experiments have been made with slow-moving currents of air irradiated with certain doses of ultra-violet, while at the same time parallel tests were carried out without irradiation. It was thereby possible to determine the dying-off rate as a function of the dose of irradiation for different micro-organisms introduced into the air by atomization of a suspension.

Whereas for Bacterium coli in an aqueous suspension a dose of about 10 mWsec/cm² was

factor to be considered when judging this difference, though it only partly explains it.

The fact is that when a liquid suspension is irradiated there is practically no diffraction or scattering of the rays on the surface of the bacterium, so that there is the same intensity everywhere inside the bacterium as in the surrounding liquid. In the case of a suspension atomized in the air it is to be presumed that owing to the curvature of the surface a certain lenticular action takes place along the boundary of the micro-organism, which action, however, cannot be calculated by geometrical-optical means because the dimension is only a few times that of the wavelength of the radiation used. Calculations made for the case of a spherical micro-organism indicate, while making allowance for interferential phenomena, that the intensity of radiation in the centre may be about twice that outside the organism.

Applications

In the application of ultra-violet sterilization there are two cases to be distinguished: volume sterilization and surface sterilization.

For microbiological manipulations, such as culture inoculations, it is often desirable to carry these out in germ-free surroundings, since in spite of the most rigorous sanitary measures it is not always possible to prevent germs being carried along by currents of air.

Sterilization can be carried out by direct irradiation of the space or place where sterile conditions are required, but this is only practicable when human beings are seldom present, because the radiation of 2537 Å has unpleasant (and in course of time injurious) effects upon the human organism. With a sufficient dose of the radiation painful erythema arises on the human skin, while, if the eye is exposed to the rays, a very troublesome, though harmless and quickly healing, inflammation of the conjunctive membrane arises (conjunctivitis).

For these reasons, in the U.S.A., where the germicidal lamp was first introduced, regulations have been laid down specifying the conditions that have to be complied with for safeguarding people in the vicinity of these lamps. Experience teaches that if these regulations are observed there is no risk of any harmful or troublesome consequences arising from the use of these lamps. It is specified, for instance, that the human skin must not be exposed longer than 7 hours per day to an intensity of at most 0.5 $\mu W/cm^2$ (for the wavelength 2537 Å). In cases of continuous irradiation (e.g. in hospital wards) the intensity on the skin must not be more than 0.1 $\mu W/cm^2$. It is the task of the designer of an installation employing germicidal lamps to make this action as effective as possible while duly observing the safety requirements.

In places where people are regularly present the lamps can be so mounted that only the upper part of the room is irradiated. The natural circulation of the air then ensures that all germs come to pass through an area of intense radiation. The lamps can be mounted in aluminium fixtures directing the rays upward, but care has to be taken that the ceiling does not reflect too much of the ultra-violet radiation, so as to avoid trouble from erythema and conjunctivitis.

As an example of indirect irradiation a case can be mentioned where a laboratory 15 m^2 in area and 4 m high had to be kept free of germs. On one of the walls at a height of 2 m a fixture was mounted with one TUV lamp. The result of the irradiation was investigated by determining the bacteria

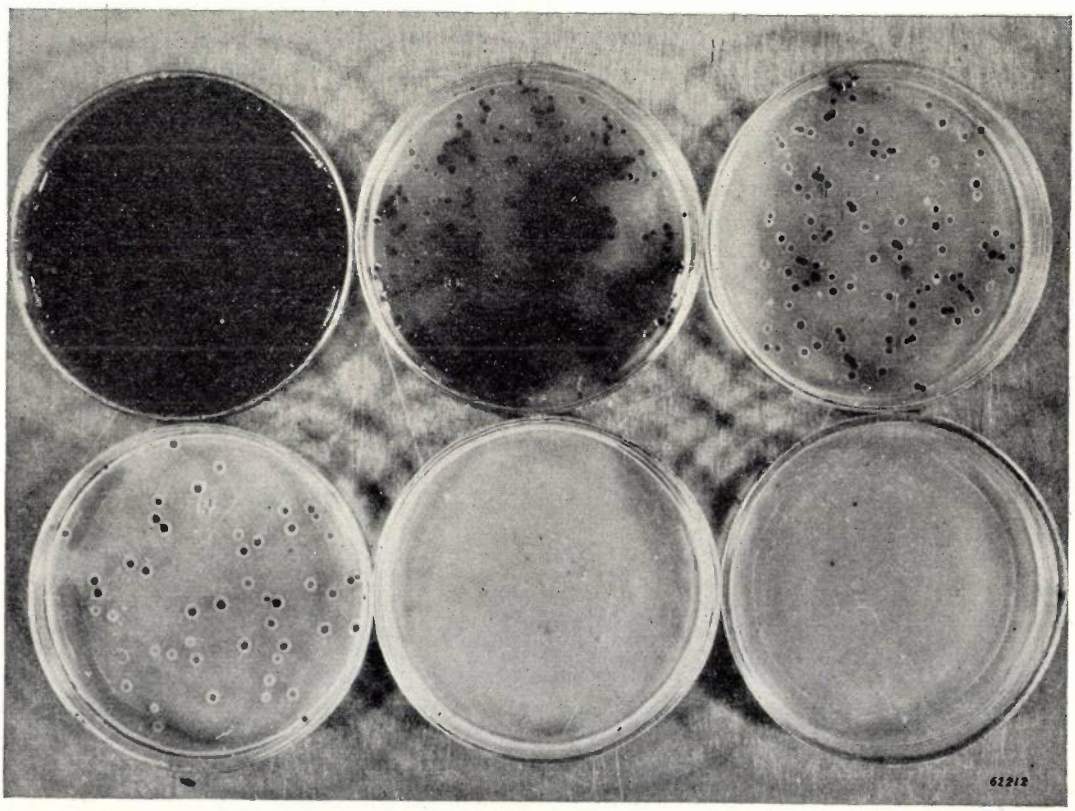


Fig. 10. Development of bacteria due to infection from the air. *Top row*: in a space where no lamp was burning, *bottom row*: germicidal lamp switched on immediately after atomization of bacteria in the air. From left to right: 2, 5 and 15 minutes after atomization.

content of samples of the air. After having atomized a culture of coli bacteria a certain quantity of air was collected with the aid of an apparatus whereby the air was blown onto a culture medium.

The result can be seen in *fig. 10*. The upper row gives an impression of the number of bacteria collected during the first 2, 5 and 15 minutes after atomizing and before the lamp was switched on.

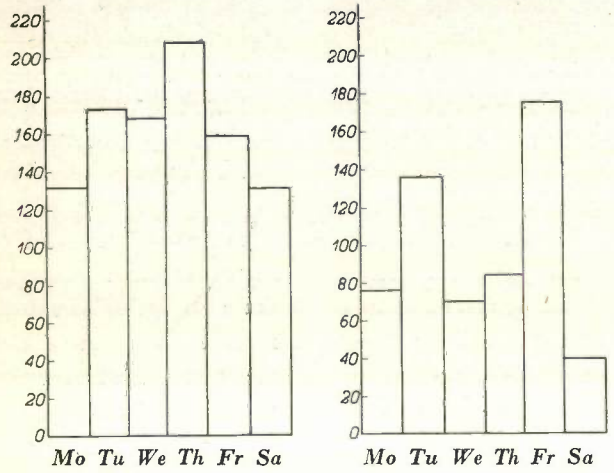


Fig. 11. Number of germs developed on a culture medium in Petri dishes in a busy, small office. On the left without TUV lamp, on the right with TUV lamp burning. From left to right the six working days in a week.

In the case of the second series of collections (lower row) the lamp was switched on immediately after the atomization, and it is seen that after 5 minutes no more bacteria were collected.

This result is not always to be expected, because as a rule bacteria from a young culture as used for this experiment have less resistance than the saprophytic micro-organisms present in large numbers in the air, and which are usually harmless. Pathogenic (disease-producing) bacteria and germs from scattered drops in the air are generally very soon killed. *Fig. 11* shows the result of irradiation in an office where there was a great deal of moving to and fro and where saprophytes were contained in the air. Several times a day for a fortnight the number of germs (bacteria and moulds) was determined that were collected in 5 minutes on a series of Petri dishes.

The graph on the left of fig. 11 represents the situation without the lamp burning, whilst that on the right shows the results with the lamp burning. There were still rather a large quantity of bacteria collected while the lamp was burning, but it may be taken that mainly the pathogenic bacteria had disappeared, so that the real effect is more satisfactory than would at first appear from the graph.

Often a more local sterilization of a certain area is required, as for instance on conveyor belts, in filling machines, washing machines, yeast vats or for inoculating cabinets in bacteriological laboratories, of which last-mentioned application a picture is given in *fig. 12*.

In the dairy industry previously sterilized tins are often transported over a rather long distance

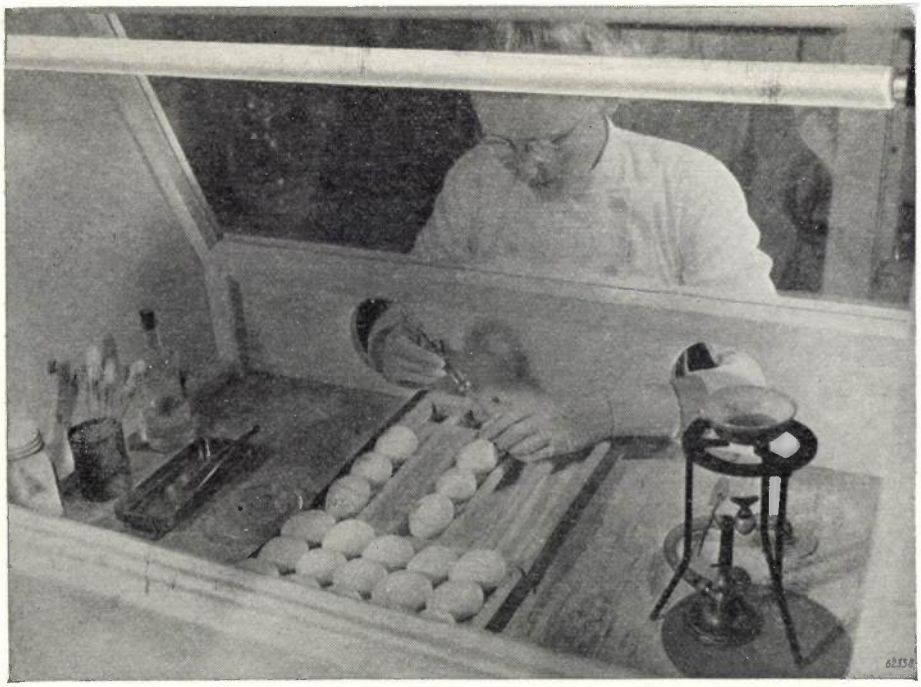


Fig. 12. Inoculating eggs with a highly infectious material, as carried out in the Institute for Tropical Hygiene at Amsterdam. The germicidal lamp is of particular importance here for safeguarding the laboratory workers.

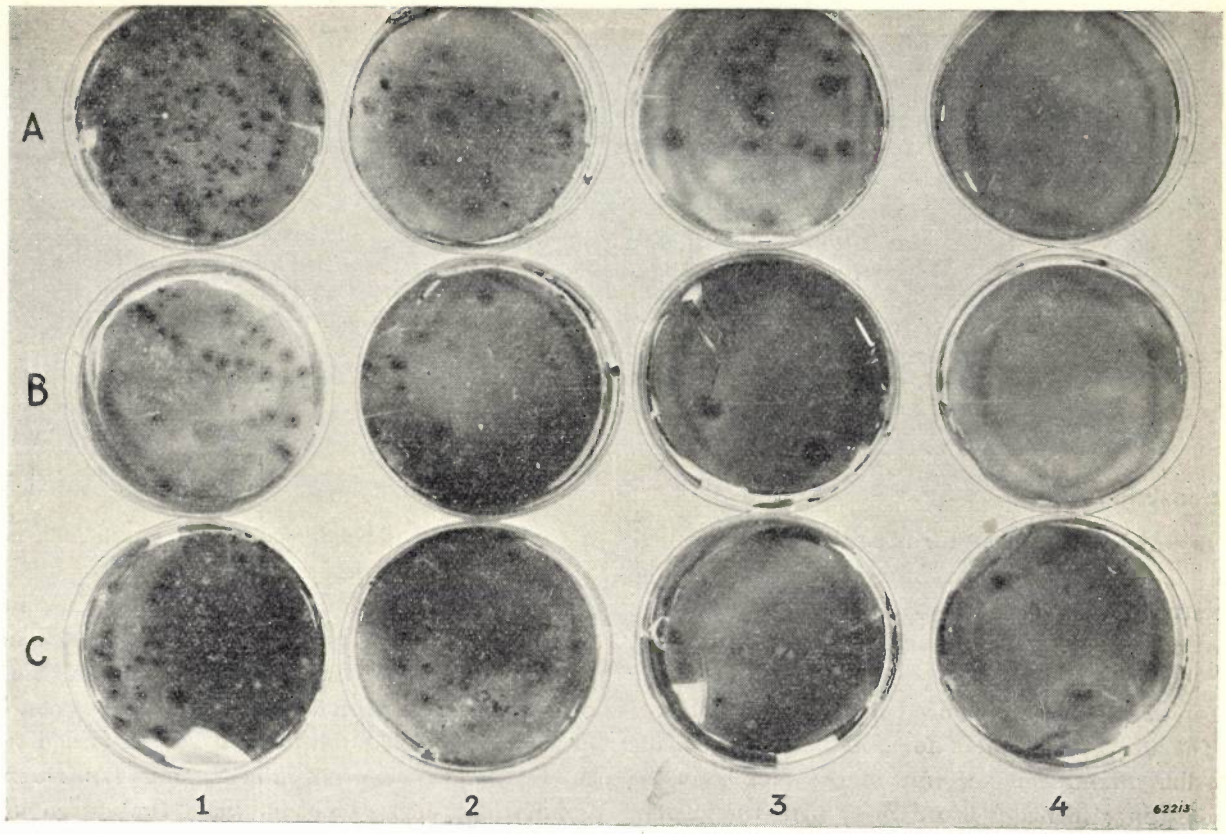


Fig. 13. Occurrence of bacteria and moulds on floors. *A* = covered with linoleum, *B* = tiled floor, *C* = parquet floor. *1* without cleansing, *2* irradiated, *3* after 3 minutes' vacuum cleaning, *4* irradiated after vacuum cleaning.

to the filling machines via conveyor belts, with the chance of infection from bacteria on the way. Here the solution is ready to hand by mounting a series of TUV lamps over the conveyor belt, so as to keep the tins sterile.

In laboratories an inoculating cabinet with a TUV lamp mounted inside it renders good service. As a special application may also be mentioned the use of the TUV lamp for preventing the growth of algae in hot-houses [3]).

Direct irradiation of surfaces has not always the desired result owing to the presence of dust, since in dust spores of bacteria and moulds are often surrounded by organic material, which intercepts the rays.

The effect can be investigated by using filter papers soaked in a suitable culture solution, say agar, pasted under sterile conditions on the surfaces to be tested. A large proportion of the germs present then stick to the agar and after "culturing" an idea can be formed of the vegetation on the paper, as shown in *fig. 13*. The first column on the left are specimens taken from a dusty linoleum-covered floor, a tiled floor and a parquet floor, while those in the second column show the result after direct irradiation of the same floors.

In practice the usual steps, like washing down, chemically disinfecting and vacuum cleaning, are taken. The effect of 3 minutes' vacuum cleaning is shown in the third column, where an appreciable reduction of the number of germs is already noticeable. It is therefore certainly advisable to make good use of a vacuum cleaner. In several cases it is warranted to follow this up with ultra-violet irradiation, as for instance over the floors of corridors in hospitals, of shower-bath cells, of doctors' waiting rooms, of polyclinics, where according to the manner in which it is applied ultra-violet irradiation may considerably reduce the number of germs. The fourth column of specimens in fig. 13 shows that then one may speak of almost perfect sterility.

Summary. Ultra-violet radiation of the wavelength 2537 Å has an "inactivating" action upon bacteria and moulds, the process of division being checked, so that no cell division takes place. As a source of this radiation a low-pressure mercury-discharge lamp, the TUV lamp, has been developed, with the aid of which it is possible to reduce considerably the number of germs in places where injurious consequences are to be feared. A dose of 5 mWsec/cm² is usually sufficient to render 90% of the irradiated bacteria inactive, the dose actually needed depending upon the species, the surroundings and the vital phase of the bacteria at the moment of irradiation.

When this lamp is employed for disinfecting rooms the fixtures must be of a suitable construction and mounted in such a way as to avoid harmful effects of the radiation upon persons present in the room.

[3]) R. van der Veen, A small greenhouse with artificial lighting for studying plant growth under reproducible conditions, Philips Techn. Rev. **12**, 1-5, 1950 (No. 1).

SATURATION OF FLUORESCENCE IN TELEVISION TUBES

by A. BRIL and F. A. KRÖGER.

535.376:621.385.832

The picture observed on the fluorescent screen of the cathode-ray tube in a television receiver is not, as a rule, a faithful reproduction of the original. One of the causes of this, though of less importance than others, is the fact that the light from the fluorescent screen does not continue to increase proportionately with the intensity of the electron beam but shows what may be called saturation. This phenomenon of saturation is not usually noticeable, but in certain cases it may become very troublesome; it may cause changes in the colour of the fluorescent light with changing load. Investigations have shown how this deficiency can be overcome.

Introduction

In cathode-ray tubes, for instance in oscilloscopes and television receivers, and in the case of the electron microscope use is made of the fluorescence of certain solids[1]) when bombarded with electrons. In this article we shall deal with the fluorescing substances (phosphors) used in television tubes.

The use of phosphors for this purpose is made possible mainly on account of the fact that the fluorescence increases in intensity with the current density of the electron beam. So long as the latter is not too high, the intensity of the light increases in direct proportion with it, but when the electron beam is of high density the increase in the intensity of the light is less than proportional; in other words, the efficiency of the fluorescence decreases when the load is high. Eventually, if the density of the electron beam is increased still further, there is hardly any increase at all in the intensity of the light, and one then speaks of saturation.

This deviation from the proportionality between brightness and current density is one of the reasons why the intensities in the television picture are not proportional to the corresponding intensities in the original. This need not be troublesome. In fact also in an ordinary photographic print the intensities of the picture are not proportional to those of the original: in this case, too, "saturation" occurs, due — inter alia — to the simple fact that the brightness of the picture can never be greater than that of the unexposed white paper.

Disproportionality tends to become troublesome, however, when a fluorescent screen is used which, in order to produce a white picture, is composed of a mixture of two phosphors (say, one with yellow and one with blue fluorescence) which saturate at different current densities. When the light is of high intensity a change in colour then arises, which of course is undesirable.

When studying this complex of phenomena, which, for the sake of brevity, we shall refer to as "saturation", we have to distinguish between the case of continuous electron irradiation, where the density of current on all points of the screen is constant in time, and the case of discontinuous (scanning) irradiation, such as occurs when the screen of a television tube is scanned by a narrow beam of electrons tracing out a raster. Assuming, for the sake of simplicity, that the cross section of the electron beam is a square, and that this beam is made to scan the whole of the surface of the image in precisely contiguous strips, then the time average of the density of current is of the same value, $\overline{J}_d$, at all points of the screen, and this value corresponds to the current density in a continuous beam, J_c, which, given an equal total current, would cover the whole area of the screen. So long as $\overline{J}_d$ and J_c are not excessive, the brightness of the screen under continuous irradiation will then be equal to that under intermittent irradiation. The actual current density, J, in the case of intermittent irradiation however is many times greater than $\overline{J}_d$. If f is the image frequency and T the time the electron beam takes to scan a certain spot, i.e. the time taken for the beam to pass through a certain point on the screen, then

$$\overline{J}_d = f J T;$$

in most television receivers $f = 25$ or 30 sec^{-1}, whilst T is in the order of 10^{-7} sec (see appendix). Thus J is a factor of about 3 to 4 times 10^5 greater than $\overline{J}_d$.

When studying this problem of saturation we shall be mainly concerned with the case of discontinuous irradiation, since this is most commonly

1) The use of fluorescing substances has been discussed at length in an article by F. A. Kröger, Applications of luminescent substances, Philips Techn. Rev. 9, 215-221, 1947.

met in practice. There are various reasons, however, why we should also consider the case of continuous irradiation. In the first place this is a theoretically simpler case and secondly continuous irradiation lends itself better to accurate experimentation. It must be borne in mind, however, that while two substances may behave exactly alike under scanning conditions, in the case of continuous irradiation they sometimes behave quite differently. An example of this is illustrated in *fig. 1*. We shall see that this is due to rather different factors prevailing in the two cases.

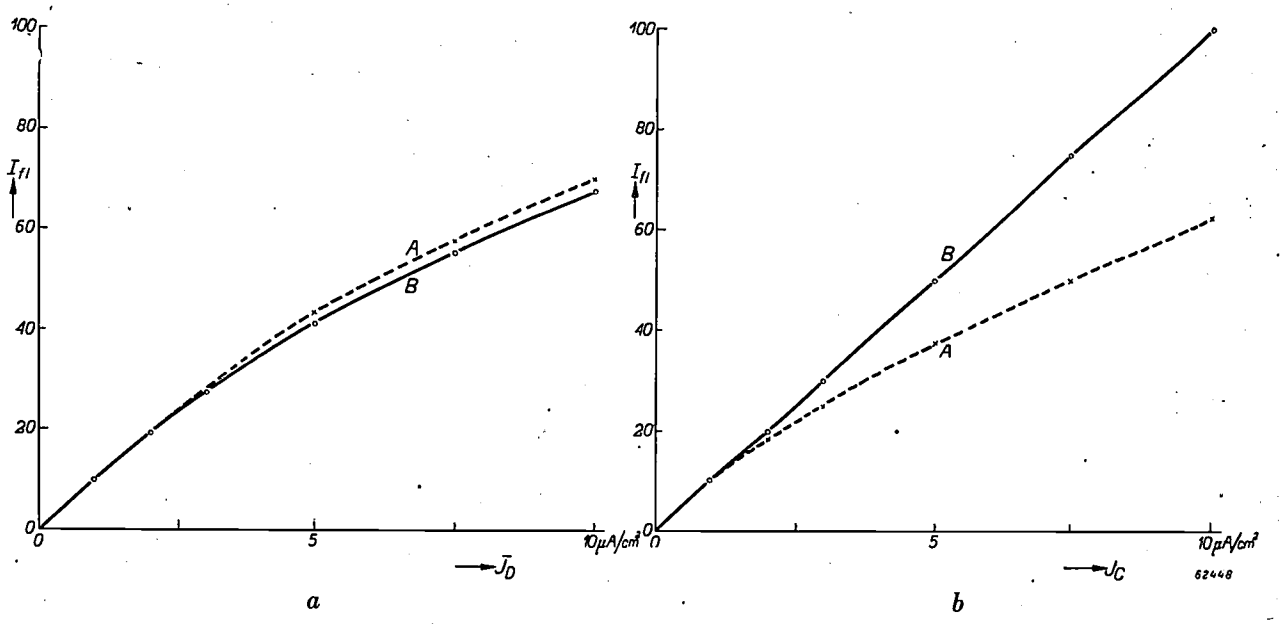


Fig. 1. *a*) Light output as function of the average current density for a zinc-silicate phosphor activated with manganese (*A*, Zn_2SiO_4 - 0,015% Mn_2SiO_4) and for a zinc-sulphide phosphor activated with silver (*B*, ZnS - 0.010% Ag), in the case of discontinuous irradiation. The two phosphors have about the same activator content and are saturated at the same average current density. *b*) Light output as function of the current density for the same phosphors as in (*a*) but for the case of continuous irradiation: it appears that the sulphide phosphor in this case does not reach saturation, whereas the silicate phosphor becomes saturated at the same current density as in the case of discontinuous irradiation. The light output has been plotted in arbitrary units, as applies also for the analogous following graphs.

Cause of saturation

The proportionality between the light output and the intensity of the electron beam is disturbed by an effect intimately related to the mechanism of fluorescence. This mechanism may be described as follows. When an electron travelling at a high velocity impinges on a crystal the energy of that electron is transmitted to a number of electrons in the crystal lattice, thereby exciting them to a state of greater energy. In most crystals the electrons excited in this way return to their original state without any emission, the excitation energy being converted into lattice energy, i.e. heat.

It is only in particular systems that the excitation energy is converted, at least in part, into fluorescence. In the case of most of these substances this is due to the presence of foreign atoms or groups of atoms (in small concentration) called activators. The excitation energy is partly transmitted to these activators, so that they in turn become excited. Sooner or later the activator centres resume their original state, this being accompanied by the emission of fluorescent light.

That part of the energy of the lattice electrons that is not transmitted to the activators is lost in the form of heat. Consequently the energy efficiency of the fluorescence must be less than 100%. The highest value reached in practice is about 20%.

Following this line of thought it can readily be understood that saturation is likely to take place. The greater the density of current, the more activators are excited. Now once an activator centre has been excited it cannot for the moment absorb more energy, since it requires some time to radiate that energy and return to its original state; the

total number of centres present being naturally limited, at high currents there will be fewer centres available to which energy can be transmitted. Thus the energy of the excited lattice electrons is converted with lower efficiency into fluorescent light than is the case with lower current density.

There is another phenomenon affecting the proportionality between current density and light output: owing to bombardment by the electrons the fluorescent screen becomes heated. Now above a certain critical temperature the fluorescence of solids tends to decrease rapidly with rising temperature [2]). When that critical level is exceeded the light output does not increase with the current at the same rate as it does in the case of lower temperatures. This phenomenon can be avoided by choosing a phosphor that has the highest possible critical temperature. We cannot go into this further within the scope of the present article.

Theory of the saturation of fluorescence

In order to obtain an insight into the factors determining this saturation let us consider the energy levels of a fluorescent crystal such as are represented diagrammatically in *fig. 2* [3]).

The basic level is indicated by *0*, an excited level by *1*. Given a certain current density J of the electron beam, a number, I, of lattice electrons in the crystal are excited per second per unit volume. Except for extremely high currents, with which we are not concerned here, I is always directly proportional to J. The system may return from the level *1* to the level *0*, without radiation taking place; let the probability of this transition in the interval of time dt be equal to αdt [4]). It may also happen that the system passes from the level *1* to a state *2* where part of the energy is transmitted to an activator centre and the latter thus becomes excited. The probability of such a process taking place in the interval of time dt is proportional to the number of activator centres not excited and thus can be represented by $\beta(N-n_2)dt$, where N denotes the total number of activator centres per volume unit, n_2 the number of excited centres per volume unit (thus $n_2 < N$) and β is a proportionality constant. The case where an activator centre is directly excited by an impinging electron is very rare and may be ignored. An excited centre returns to its original state with the emission of fluorescent light; let the probability of this transition taking place in the interval of time dt be γdt.

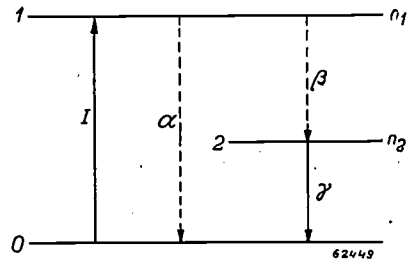


Fig. 2. Energy levels in a fluorescing crystal; *0* initial level, *1* excited level of the lattice, *2* state at which an activator is excited. The quantities α, β and γ determine the probabilities of the various transitions in the crystal. The occupation of the levels *1* and *2* is denoted by n_1 and n_2 respectively. Transitions without radiation are indicated by dotted lines.

The transition probabilities for the three processes mentioned are thus α, $\beta(N-n_2)$ — which quantity in the case of not too high a load is in the order of βN — and γ, respectively. In practice it is found that, to a close approximation, α and βN amount to at least 10^8 sec^{-1}. For the ratio of α to βN a figure greater than 3 is usually found. The constant γ may differ considerably from one phosphor to another; taken on the whole it lies between 10^2 and 10^6 sec^{-1}. The reciprocal value $1/\gamma$ of this constant is the average lifetime, τ, of an excited activator centre; τ thus varies between 10^{-2} and 10^{-6} sec.

Denoting the occupation (per volume unit) of the level *1* by n_1 and that of the level *2* by n_2, the change of n_1 and n_2 with time can be formulated by the differential equations:

$$\frac{dn_1}{dt} = I - \alpha n_1 - \beta n_1 (N-n_2), \quad . \quad . \quad (1)$$

$$\frac{dn_2}{dt} = \beta n_1 (N-n_2) - \gamma n_2 . \quad . \quad . \quad . \quad (2)$$

After what has already been said there will be no ambiguity about the meaning of the various terms. It only remains to add that the term γn_2 in (2) represents the number of light quanta emitted per second per unit of volume. It is therefore important to know how n_2 behaves as a function of the number of lattice electrons I excited per second and of the time t, bearing in mind the distinction that has to be made between continuous

[2]) See F. A. Kröger and W. de Groot, Influence of temperature on the fluorescence of solids, Philips Techn. Rev. **12**, 6-14, 1950 (No. 1).

[3]) For an extensive mathematical treatment of saturation see A. Bril, On the saturation of fluorescence with cathode-ray excitation, Physica, The Hague **15**, 361-379, 1949 (No. 3/4).

[4]) It is assumed that this is a "monomolecular" process, i.e., that the probability for a transition is proportional to the first power of the occupation of the level *1*. The "bimolecular" case, where we have to do with a recombination of electrons and "holes" (see the articles quoted in footnotes [2]) and [3])) and the probability of a transition taking place is proportional to the square of the occupation of level *1*, is left out of consideration here.

and discontinuous irradiation. The manner in which this behaviour is calculated is explained in the appendix. The results are as follows.

For the case of continuous irradiation, the relation between I and the number of fluorescent quanta L_c emitted per second per unit of volume is found, to a good approximation, to be:

$$L_c = \gamma n_2^* \approx \frac{\eta_0 I}{1 + \eta_0 I/\gamma N} \quad . \quad . \quad . \quad . \quad (3)$$

where n_2^* represents the constant ultimate value of n_2 when the irradiation has lasted some time, and η_0 is the quantum efficiency of the fluorescence at a low current density; i.e. the number of light quanta emitted per excited electron in the original lattice. This quantum efficiency η_0 is given by

$$\eta_0 = \frac{\beta N}{\alpha + \beta N} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Henceforth we shall assume that $\alpha/\beta N \approx 10$, so that $\eta_0 \approx 0.1$.

From (3) it is seen at once that the proportionality between L_c and I begins to be disturbed as soon as the term $\eta_0 I/\gamma N$ in the denominator — which at low load is $\ll 1$ — approaches the value 1, thus when I becomes so great as to satisfy the condition

$$\eta_0 I \approx \gamma N = N/\tau \quad . \quad . \quad . \quad . \quad (5)$$

For the case of discontinuous irradiation the intensity of the light depends also upon the time T the electron beam takes to scan a certain spot. Since T is much smaller than τ, no stationary state will be reached. For the average number of quanta $\bar{L}_d$ emitted per second we therefore find a different expression, viz., to a close approximation:

$$\bar{L}_d \approx fN\,(1 - e^{-\eta_0 IT/N}), \quad . \quad . \quad . \quad (6)$$

where $f =$ the image frequency. Introducing the time average of the number of excited electrons $\bar{I} = fIT$ we get the formula

$$\bar{L}_d \approx fN\,(1 - e^{-\eta_0 \bar{I}/fN}) \quad . \quad . \quad . \quad . \quad (6a)$$

In this case there will be a deviation from the proportionality between $\bar{L}_d$ and $\bar{I}$ as soon as $\eta_0 \bar{I}/fN = \eta_0 IT/N \approx 1$, thus when

$$\eta_0 \bar{I} \approx fN \quad \text{or} \quad \eta_0 I \approx N/T, \quad . \quad . \quad . \quad (7)$$

which result may be compared with formula (5).

In the above results *no* account has been taken of the effect of quenching centres, i.e. centres of the same nature as activator centres but having the property of returning to the original state from the excited state without emission. For the purpose of our investigation this omission is allowed because we are assuming for the time being that the concentration of activator centres is much greater than that of quenching centres. We shall revert to the quenchers later.

Comparison between continuous and intermittent irradiation

To compare the saturation taking place with one and the same phosphor in the case of continuous irradiation with that in the case of discontinuous irradiation the light intensity can be plotted as a function of the current density in the one case and as a function of the average current density in the other. The curves thus produced are called characteristics. From equation (3) it follows that with continuous irradiation the number of light quanta emitted per unit of volume per second approaches γN if the current density is allowed

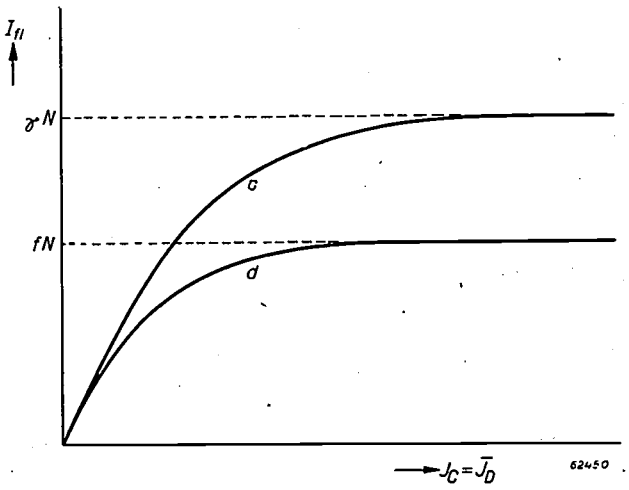


Fig. 3. Comparison of the characteristics of one and the same phosphor under continuous irradiation, curve *c*, and under discontinuous irradiation, curve *d*. Plotted on the same scale as abscissa is the current density and the average current density in the beam respectively.

to increase indefinitely. With intermittent irradiation it follows from (6a) that the limit for the number of light quanta in this case is equal to fN (assuming f to be smaller than γ). For small values of the current density the number of quanta emitted is equal (see above). When the two characteristics are reproduced in one graph (*fig. 3*) we thus find two curves which have the same initial slope

but a different "ceiling", namely γN for the case of continuous and fN for the case of scanning irradiation.

The image frequency f in British television receivers is equal to 25 sec^{-1}; as we have already seen, γ may vary between 10^2 and 10^6 sec^{-1}. A phosphor with a high γ-value (short lifetime of the excited centres) will behave quite differently from one with a low γ-value. If the quotient γ/f is in the order of unity, as is the case with the lowest values of γ, such as may occur with silicate phosphors activated with manganese, then the "ceiling" is of about equal height for both curves and, hence, if under scanning conditions saturation becomes noticeable with a certain $\overline{J}_d$, then under continuous irradiation saturation will likewise occur with an approximately equal value of J_c. But when $\gamma \gg f$, as is the case with many sulphides, for instance those activated with silver ($\gamma \approx 10^5$), the ceiling for continuous irradiation may be a factor 1000 higher than that for scanning conditions. In that case, in the range of current densities where saturation becomes noticeable under intermittent irradiation, there will be no sign of any deviation from the proportionality under continuous irradiation. To carry out experiments regarding saturation also for the case of continuous irradiation a 1000 times greater current density would be needed, which is not practicable.

Comparison of the characteristics of two phosphors

In order to compare the fluorescence of two different phosphors, e.g. under scanning conditions, the two characteristics can again be plotted in one graph. To be able to judge at once any difference in saturation, the initial slopes of the two curves can be made to coincide by enlarging the scale for the ordinate of one curve by a factor equal to the ratio of the initial efficiencies of the two phosphors. The curve which, after this transformation, has the lowest ceiling corresponds to the phosphor showing greatest saturation.

For discontinuous irradiation this saturation is governed entirely by fN, and thus, since with a given television system f is a constant, by N. Phosphors with a high N-value (large activator content) will therefore have less tendency towards saturation than those with a small N-value. In its generality this conclusion holds equally for a comparison of two phosphors under continuous irradiation, but in this case the saturation is governed by the factor γN, thus being dependent upon γ in addition to N; two phosphors with about the same activator content but with different γ-value will thus differ in saturation, that having the greater γ-value showing the less saturation.

Experimental results

Following upon the theoretical considerations derived in the foregoing, we shall now discuss some experimental results. First of all we must account for the difference in behaviour of the two phosphors in fig. 1. From the fact that the two curves in fig. 1*a* practically coincide it follows that for the two phosphors in question the number of fluorescent centres per volume unit must be about equal (still disregarding the effect of quenchers), whilst the marked difference between the curves in fig. 1*b* points to a considerable difference in the value of γ. And indeed the value of τ for the silicate phosphor A is about 10^{-2} sec, whilst the lifetime of the excited centres in the case of the sulphide phosphor B is about 2×10^{-5} sec. Consequently A will very soon show saturation under continuous irradiation, whereas with the same current B will show practically no saturation at all.

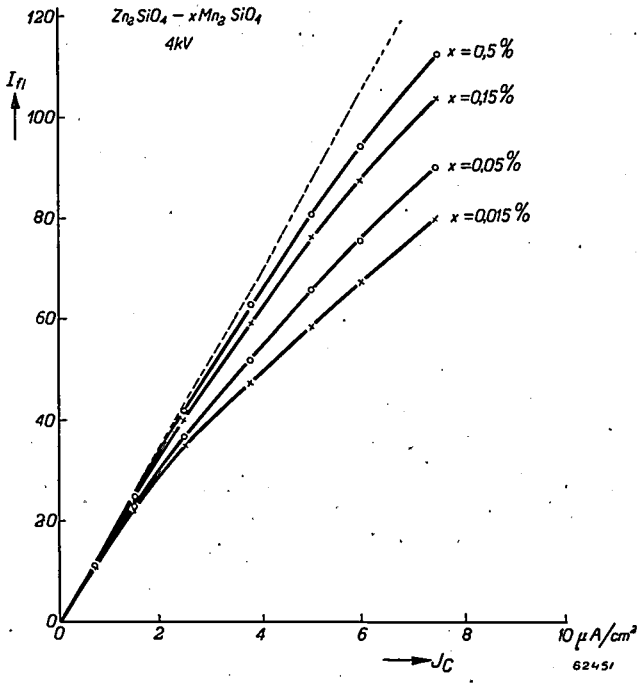


Fig. 4. Characteristics for continuous irradiation of a number of zinc-silicate phosphors activated with manganese. Saturation diminishes with increasing activator content. The characteristics have been brought to the same initial slope.

Since the value of γ ($\approx 10^2$) for A is comparable to that of f ($= 25$) the saturation of A under continuous irradiation will become noticeable with about the same current density as in the case of discontin-

uous irradiation (plotting the intensity of light against the average current density $\bar{J}_d$ and comparing the curves for the same number of $\mu A/cm^2$).

The influence of the concentration of activators, thus of N, upon saturation is clearly seen from *fig. 4*, where the characteristics have been plotted for a number of silicates activated with different quantities of manganese. Since γ is practically independent of N, here N is the only variable. It is seen that the characteristic of the phosphor indeed becomes less curved as the concentration of activators increases (increasing manganese content).

For application in television cathode-ray tubes a number of phosphors are available to choose from. The most important are the zinc and zinc-cadmium sulphides and sulphoselenides activated with silver, and the zinc, zinc-beryllium and calcium-magnesium silicates activated respectively with manganese and titanium. The sulphides have a small activator content (about 0.01%) but a high efficiency (about 15 to 20%), whereas the silicates have a much larger activator content (0,1 to 2%) but a lower efficiency (about 7%). For applications where saturation is not likely to occur, as in the case of direct-vision television tubes, sulphides will therefore be mostly used. In projection-television tubes — these are small tubes with very high luminosity of the picture which can be magnified for projection onto a screen[5]) — the load is so heavy that the avoidance of saturation is more important than a high efficiency at a small current density, and it is for this reason therefore that silicates are used in these tubes.

Mutual matching of the characteristics of the components of a mixed phosphor

The white colour of the fluorescence in television tubes is usually obtained with the aid of a mixture of two phosphors, one with a blue and the other with a yellowish-green fluorescence. It is then necessary that under scanning conditions the colour of the screen remains the same for any current density, which means to say that the ratio of the intensity of the blue light to that of the yellow light must be practically constant. This implies that when the initial slopes of the two characteristics are made equal the curves must exactly coincide one with the other. If this condition is satisfied the characteristics are said to be uniform. A mixture of two phosphors not having uniform characteristics will show a fluorescence varying in colour with the intensity of the irradiation as soon as the latter becomes so great that one of the phosphors begins to show saturation. There is not much risk of this occurring with a silicate mixture, which has little tendency towards saturation, but it must certainly be taken into account in the case of a sulphide mixture or a mixture of a sulphide and a silicate.

Let us suppose that we have a white-fluorescing mixture of phosphors composed of a blue-fluorescing zinc sulphide activated with silver and a yellow-fluorescing zinc-cadmium sulphide activated with silver. The "yellow" phosphor usually tends towards saturation at a smaller average current density than the "blue" one. How, then, can the shape of one of the characteristics be so altered as to reach uniformity? This can be done in three ways:

1) Increasing the concentration of activators in the yellow component, thus raising the ceiling of the yellow characteristic to a higher level. There is a limit, however, to which one can go in this direction, because the efficiency of the phosphor reaches a maximum at a certain concentration of activators. Moreover, the solubility of the activator atoms in the crystal lattice is limited. Therefore not very much can be achieved by this method, in a general way of speaking.

2) Reducing the concentration of activators in the blue component, thus lowering the ceiling of the blue characteristic. Reduction of the activator concentration has scarcely any influence on the efficiency of fluorescence under low load. It is only with the very smallest activator concentrations that the initial efficiency is lowered, and then either the concentration is so small as to cause saturation to occur already with low intensities of the irradiation, or else it becomes of the same order as the concentration of the quenching centres, which as a rule are present only in a small concentration. The following table shows the effect of the activator content on the initial efficiency for the case of a zinc-sulphide phosphor activated with silver (voltage across the cathode-ray tube 10 kV).

Activator content in atomic percent.	Initial efficiency in percentage
1×10^{-2}	22
6×10^{-3}	22
4×10^{-3}	22
1×10^{-3}	18

In *fig. 5* an example is given of the matching of two characteristics by the method of reducing the activator concentration in the less saturating component.

[5]) Cf J. de Gier, Projection-television receiver, II. The cathode ray tube, Philips Techn. Rev. **10**, 97-104, 1948.

When applying this matching method care has to be taken to avoid the occurrence of self-activation, by which is meant counteraction of the reduction of the activator concentration by formation of new activator centres by the atoms of the crystal lattice itself; when this arises there is hardly any change in the total concentration of activators. Self-activation can be avoided by applying a special method of preparation.

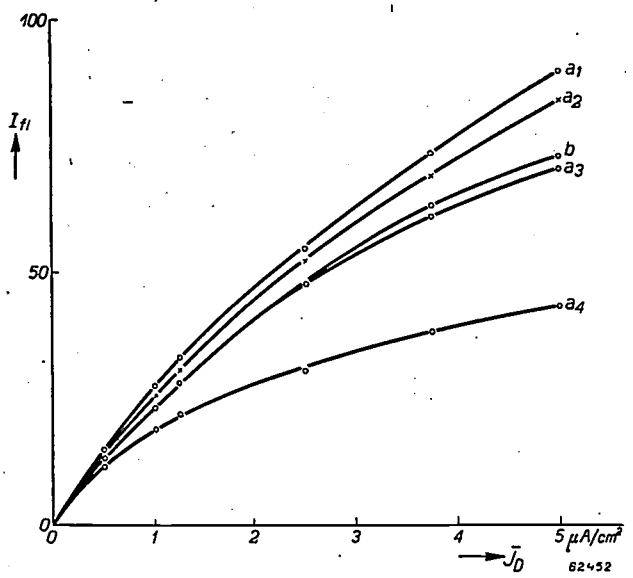


Fig. 5. Matching characteristics by reducing the activator content in the less saturating component. a_1, a_2, a_3 and a_4: characteristics of ZnS-xAg (blue) in which x = 0.01, 0.006, 0.004 and 0.001 atomic percent; b = characteristic of (Zn,Cd)S-Ag (yellow).
Matching is best when $x \approx 0.004\%$. The curves have been brought to the same initial slope.

3) Adding extra quenching centres in the lattice of the yellow component. A considerable portion of the energy otherwise almost entirely taken up by the activators when there are few quenching centres is then absorbed by the quenchers and as a result the intensity of the fluorescent light is reduced; thus the efficiency η_0 of the fluorescence is lowered. With intensities of irradiation leading to saturation in the absence of quenchers there will then be correspondingly less tendency towards saturation. This all finds expression in a flatter but straighter curve of the characteristic. However, even with the addition of quenchers, saturation ultimately sets in when a certain intensity of irradiation is reached, since the ceiling of the characteristic of course remains the same, namely fN.

Obviously this addition of quenchers does more harm than good. The efficiency at low intensities of irradiation usually drops so much that this method of matching the characteristics will only be applied in cases where no more than a slight correction is desired. An example of what can be achieved with this method is given in *fig. 6*; by the addition of 0.005% nickel to the yellow-fluorescing phosphor the efficiency at low load is reduced from 15% to 3%. It goes without saying that in order to maintain the white colour of the blended light, relatively more of the "yellow" phosphor has to be added.

Change of colour with defocusing of the electron beam in television tubes

Finally a phenomenon which may be noticed in the case of discontinuous irradiation of the television screen must be discussed. When the characteristics of the two components of a white-fluorescing mixture are uniform and thus under high load both components are equally saturated and no change takes place in the colour, colour-changes may still occur when the electron beam is defocused. Such is the case, for instance, when a blue-fluorescing calcium-magnesium silicate activated with titanium is mixed with a yellow-fluorescing zinc-beryllium silicate activated with manganese: when the beam is defocused while scanning the screen under high load the colour assumes a more bluish tint. At first sight one would be inclined to attribute this to an earlier saturation of the blue component, but since the characteristics are uniform this cannot be the cause. The explanation is as follows.

As already observed, we are dealing here with a case where the irradiation is discontinuous. Assuming that with the beam properly focused the lines of the raster do not partly overlap, then a certain spot on the screen is irradiated once every 0.04 sec. When the beam is defocused the lines will overlap, with the result that instead of the electron beam striking a certain spot once in each complete scanning it strikes that spot several times in succession at intervals equal to the time in which the beam traces a line of the raster, in our case about 10^{-4} sec.

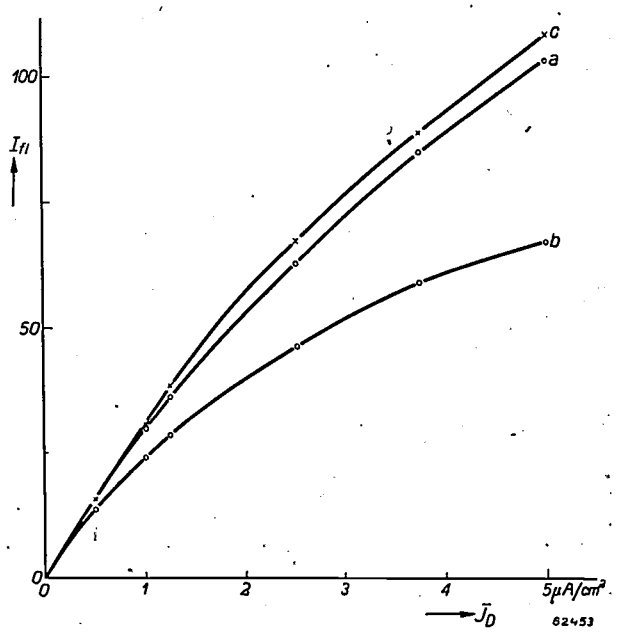


Fig. 6. Matching of characteristics by adding quenchers. The characteristic a of ZnS-0.01% Ag (blue) coincides almost with curve c, the characteristic of (Zn,Cd)S-0.01%Ag-0.005% Ni (yellow). In the absence of nickel as quencher in the latter phosphor the yellow characteristic, curve b, deviates considerably from the blue one. All curves have been brought to the same initial slope.

Since the blue component of the phosphor mixture has a τ-value of 2.5×10^{-5} sec, practically all the activator centres of this phosphor will again be in their original state each time the defocused beam crosses the spot considered, as represented in *fig. 7a*. The τ-value of the yellow component, however, is about 0.006 sec, which means that in 10^{-4} sec the number of activator centres excited by the previous passage of the electron beam is not appreciably reduced. Such is illustrated in fig. 7*b*. The consequence of the difference in τ will be that, when the electron beam reaches the intensity at which, with proper focusing, both components show saturation, in the case of defocusing of the beam the saturation of the blue component will be reduced but not, or scarcely so, that of the yellow component. As a result the colour of the fluorescence assumes a more bluish tint.

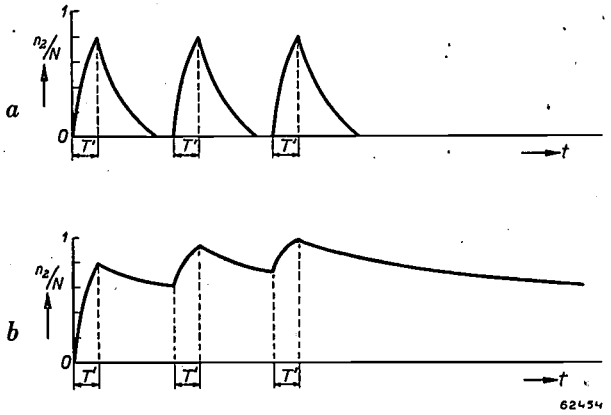


Fig. 7. When the electron beam in a television tube is slightly defocused a certain point on the fluorescent screen is irradiated not once during the scanning of the whole of the screen but several times in succession, each time a new line is traced. The variation of the number of excited activator levels as a function of time is indicated, for the case of three successive passages during a scanning, *a*) when the average lifetime of the excited levels is shorter than the interval of time between two successive passages (in our case about 10^{-4} sec), *b*) when it is longer than that interval. T' is the time the widened electron beam takes to scan a certain spot of the screen.

This can be confirmed by taking a look at the fluorescent screen under a high load through filters passing either blue or yellow light. When a blue filter is used the intensity of the light diminishes considerably during focusing, whereas with a yellow filter the reduction is only small. The fact that some reduction does indeed take place is due to the interlacing of the raster in the case of television tubes; owing to two adjaccent lines not being traced in one single scanning, part of the defocused beam may still strike yellow activator centres that have not been excited. (Of course this applies only when there is not too much defocusing.)

The foregoing phenomenon would seem to be contradictory to our earlier statement that in the case of intermittent irradiation the saturation is independent of τ. In our explanations concerning this case, however, it was assumed that the time elapsing between two successive passages of the electron beam is long compared with the duration of this decay time, a condition which is not complied with in the above example for the yellow component.

Appendix: Derivation of the formulae (3) and (6)

When a phosphor is subjected to a continuous electron bombardment in course of time a stationary state will be eached; there is then no further change in the occupation of the levels and n_1 and n_2 assume constant values which we shall denote by n_1^* and n_2^*. Since the differential quotients in the left-hand members of (1) and (2) may be taken as equal to zero, we have

$$I - \alpha n_1^* - \beta n_1^* (N - n_2^*) = 0 , \quad . . . . . \quad (8)$$

$$\beta n_1^* (N - n_2^*) - \gamma n_2^* \quad = 0 . \quad . . . . . \quad (9)$$

For the number of fluorescence quanta emitted per second and per unit of volume we find:

$$L_c = \gamma n_2^* = \frac{I}{\frac{\alpha + \beta N}{\beta N} + \frac{\alpha}{\beta N} \frac{n_2^*}{N - n_2^*}} . \quad . . . . \quad (10)$$

The ratio L_c/I represents the quantum efficiency, i.e. the number of quanta emitted per excited electron. With a small current density, thus small n_2^*, this efficiency is

$$\eta_0 = \frac{\beta N}{\alpha + \beta N},$$

with which formula (4) has been derived. As already stated, it is taken that $\alpha/\beta N \approx 10$; by ignoring a negligible quantity in (10) we find formula (3):

$$L_c = \frac{\eta_0 I}{1 + \eta_0 I/\gamma N} .$$

In the case of discontinuous irradiation, as occurs in television tubes, such a simple mode of reasoning is not possible. Let us suppose, for the sake of simplicity, that the scanned part of the screen is a square. The picture is divided into m contiguous strips. If the cross section of the beam is a square the whole area of the picture will therefore consist of m^2 picture elements of the size of that cross section. Per second the beam passes over fm^2 picture elements (f being again the image frequency), so that the time the beam takes to scan a certain spot of the screen is $T = 1/fm^2$ sec. If, for example, $m = 625$ and $f = 25$ sec^{-1}, then $T \approx 10^{-7}$ sec.

It is obvious that in this interval of time, which is very short compared with the lifetime τ of the excited activator centres, no stationary state is reached in the irradiated phosphor. In order to ascertain the behaviour of the phosphor we must therefore first consider the variation of n_2 with time. In principle this is possible by eliminating n_1 from (1) and (2), but the result cannot be written in a concrete form.

The solution of the equations (1) and (2), however, can be appreciably simplified, without seriously affecting the correctness of the conclusions, by ignoring in (1) the quantity n_2 with respect to N. For n_1 we then find

$$\frac{dn_1}{dt} = I - (\alpha + \beta N) n_1 ,$$

with as solution:

$$n_1 = \frac{I}{\alpha + \beta N} (1 - e^{-(\alpha + \beta N) t}) .$$

Thus it is seen that n_1 assumes a stationary value in a very short time ($\approx 10^{-8}$ sec). (Considering that after this short time level *2* will not by any means be filled, the fact of n_2 being ignored in equation (1) will not have been of any great consequence.) Introducing the stationary value in (2), then, in view of (4), we get:

$$\frac{dn_2}{dt} = I\eta_0 - \left(\frac{I\eta_0}{N} + \gamma\right) n_2 , \quad . . . . \quad (11)$$

the solution of which is:

$$n_2 = \frac{I\eta_0}{\frac{I\eta_0}{N}+\gamma}\left[1-e^{-\left(\frac{I\eta_0}{N}+\gamma\right)t}\right] \quad . \quad . \quad . \quad . \quad (12)$$

Thus we have found the relation between n_2 and t.

We now have to find the number of light quanta emitted by the phosphor as an average per second and per unit of volume.

When the beam has passed the element of the screen considered, thus at the moment $t = T$, the number of excited centres per unit of volume is equal to $n_2(T)$. Sooner or later, on an average after an interval of time τ, all these centres will emit a quantum. The number of light quanta per current impulse can thus simply be taken as equal to $n_2(T)$, and the average number of light quanta per second is equal to

$$\overline{L}_d = f n_2(T).$$

(Actually the total number of quanta emitted per impulse is

$$n_2(T) + \gamma \int_0^T n_2 dt.$$

The last term represents the number of quanta already emitted in the interval of time $0 < t < T$. This, however, is now smaller than $\gamma T \times n_2(T)$ and, since γT is usually $\ll 1$, namely $10^{-5} < \gamma T < 0.1$, the second term may be ignored.)

Hence we find:

$$\overline{L}_d = f n_2(T) \approx f \frac{\eta_0 I}{\gamma + \eta_0 I/N}\left[1-e^{-\left(\frac{\eta_0 I}{N}+\gamma\right)T}\right].$$

Since $\gamma T \ll 1$, in view of (7) this may to a sufficient approximation be replaced by formula (6).

Summary. In the case where phosphors are brought to fluorescence by excitation with electrons, when a high current density is reached in the electron beam the efficiency of the fluorescence is found to decrease. This is due to the number of activator centres present in the crystal lattice being limited This "saturation" of the fluorescence appears, in the case of continuous bombardment with electrons, to depend both upon the activator concentration and upon the average lifetime of the excited activators. In the case of discontinuous irradiation, as takes place when the screen of a television tube is scanned with an electron beam, there is no longer any dependency upon the average lifetime of the activators if that lifetime is shorter than the interval between two successive passages and longer than the time the electron beam takes to scan a certain spot.

A comparison of the properties of silicate phosphors and sulphide phosphors shows that it is advantageous to use sulphide phosphors in tubes in which not too large current densities occur, whilst in tubes with a heavy load on the fluorescent screen silicate phosphors are to be preferred, notwithstanding the fact that as a rule their efficiency is less than that of sulphide phosphors.

In order to avoid a change in colour under a high current density when using a white-fluorescing mixture of two phosphors, the characteristics of the two components must be uniform. Various possibilities are discussed for matching the characteristics of the components one with the other. Even when a mixture with uniform characteristics is used colour-changing may still occur when the beam is defocused. This is due to the difference in the average lifetime of the activators in the components compared with the time taken for one line of the picture to be traced.

VOL. 12 No. 5, pp. 129-156 NOVEMBER 1950

Philips Technical Review

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

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

AN INSTANT-STARTING FLUORESCENT LAMP IN SERIES WITH AN INCANDESCENT LAMP

by W. ELENBAAS and T. HOLMES*).

621.327.43

To operate a fluorescent lamp it has to be connected in series with a certain impedance. Usually a choke is employed for this purpose, and not a resistance, because with the latter there would be a rather high loss of power. This objection, however, can for the greater part be met by using an incandescent lamp as stabilizing resistance, which has the advantage that this lamp also gives light. By making some changes in the construction of the fluorescent lamp for connecting it in series with an incandescent one a solution has been found which offers certain advantages and at the same time dispenses with the starter otherwise needed for ordinary fluorescent lamps.

The tubular fluorescent lamps ("TL") hitherto commonly used in series with a stabilising apparatus, or ballast, when connected to A.C. mains of 220 V have an ignition voltage which is very much higher, with cold electrodes, than the peak voltage of the mains. Even after the electrodes have been heated up to glowing temperature the voltage required for the first ignition is usually still higher than the peak of the alternating voltage available.

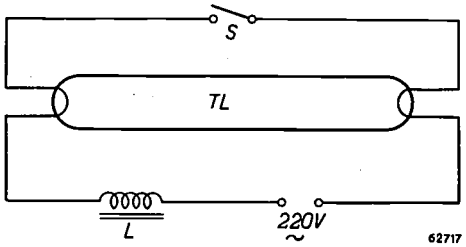


Fig. 1. Basic circuit of a "TL" lamp (*TL*) operating from A.C. mains. *L* series choke, *S* starter switch.

For the ignition of a "TL" lamp the system illustrated in *fig. 1* is often employed. When the switch *S* is closed an alternating current passes through the choke *L* and the two electrodes of the lamp, thereby heating the latter. Upon the current being interrupted by opening the switch, then, owing to the inductance of the choke, a voltage surge arises which is usually sufficient to cause the lamp to ignite. The electrodes are then kept heated by the discharge current and the re-ignition voltage remains low enough to keep the lamp burning without any further help.

The switch *S* does not necessarily have to be operated by hand but can be operated automatically. In practice automatic ignition switches (starters) are invariably used, one form of which, the glow-discharge starter, is depicted and explained in *fig. 2*. This and other forms of starters have already been fully dealt with in this journal [1][2]).

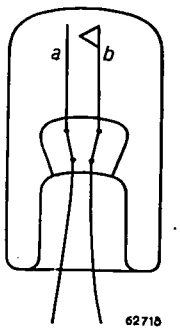


Fig. 2. Glow-discharge starter automatically performing the function of the switch *S* in fig. 1. When the mains voltage is switched on a glow discharge takes place between the electrodes *a* and *b* of the starter, thereby raising the temperature of the electrode *a*, made in the form of a bimetal, to such a level as to cause it to make contact with *b*. This stops the glow discharge, the bimetal cools down and the contact is broken. If this happens at a moment when the current is so low that the voltage surge at the choke is insufficient to cause the lamp to ignite, then the process repeats itself until the lamp does ignite. The working voltage of the "TL" lamp is not high enough to maintain a glow discharge in the starter.

*) Philips Hamilton Works, Hamilton, Scotland.

[1]) A. A. Kruithof, Tubular luminescence lamps for general illumination, Philips Techn. Rev. 6, 65-73, 1941.

[2]) Th. Hehenkamp, A rapid-action starter switch for fluorescent lamps, Philips Techn. Rev. 10, 141-149, 1948/49.

The glow-discharge starter is highly dependable and has a long life, but there is the drawback that it takes two to four seconds for the lamp to ignite after switching on. This is not a serious objection for factory and office lighting where the lamps are switched on only once or twice a day, but it is inconvenient for house lighting, and that is why attempts have been made to get starters with a more rapid action. The article quoted in footnote [2]) describes a starter that has a delay, under normal conditions, of only 0.3 to 0.4 sec.

Pros and cons of an incandescent lamp compared with a stabilizing choke

The starter, however, does not provide the solution in all cases. We have in mind here the case where it is desired to replace the choke of the "TL" lamp by a series resistance. The power consumption of such a resistance is about as great as that of the "TL" lamp itself and many times greater than the losses in a choke, but this objection can for the greater part be avoided by using an incandescent lamp as resistance, this lamp then also producing a certain amount of light. An incandescent lamp is certainly a less efficient source of light than a "TL" lamp, but the reduced efficiency is offset by the following advantages:

1. An incandescent lamp (plus holder) is much cheaper than the choke hitherto used with "TL" lamps, and also lighter; the "TL" lamp fixture can then be made lighter too.

2. The power factor of an A.C. fed "TL" lamp is approximately 1 with a series resistance but only about 0.5 with a choke. (In the case of large lighting installations, in order to get a good power factor with chokes, half the number of lamps are usually provided with a ballast consisting of a choke in series with a capacitor. The inductance and the capacitance are so chosen that the current is of just the right value and leading in phase with respect to the mains voltage by about the same amount as the current of the other lamps is lagging. Naturally such a ballast with choke and capacitor is still larger, heavier and more expensive than a choke alone.)

3. An incidental advantage is that an incandescent lamp does not give rise to any hum; the hum caused by a choke can only be eliminated at the cost of special measures.

4. With the combination of a "TL" and an incandescent lamp in the same fixture a blended light is obtained which, as regards both distribution and spectral composition, is a good compromise between fluorescent light and incandescent light alone.

5. To use fluorescent lamps in places where only D.C. mains are available there is no alternative but to use a stabilizing resistance, in which case incandescent lamps will of course preferably be employed as resistances.

Let us now consider the difficulties arising when with a normal "TL" lamp the choke on the system according to fig. 1 is replaced by a suitably dimensioned incandescent lamp.

There are two difficulties to be faced. In the first place, in order to get a reasonable amount of light from the incandescent lamp in series with the "TL" lamp provision has to be made for its filament to be heated to the normal temperature (about 2700 °K) with the r.m.s. value of the voltage occurring across it when the lamp is burning. Where an incandescent lamp is connected in series with a "TL" lamp of 40 W to A.C. mains of 220 V, the r.m.s. value is about 135 V. When, however, the starter contacts are closed, almost the full mains voltage (reduced only by the voltage loss in the two filaments of the "TL" lamp) comes to lie across the incandescent lamp, which is thus heavily overloaded when switching on.

The second difficulty lies in the fact that, owing to the absence of inductance, when the contact opens, the interruption of the current does not set up a voltage surge. The action of the starter only heats the electrodes of the "TL" lamp, and, although this causes the ignition voltage to drop, it is not reduced far enough for reliable ignition without the aid of a voltage surge.

If the ignition voltage of the "TL" lamp with the electrodes in the cold state were not higher than the peak of the A.C. voltage available, or if a sufficiently high alternating voltage were available, there would be no need to preheat the electrodes and consequently no starter would be required, which is of course in itself an advantage, and then both the difficulties mentioned would be overcome.

Now there are two methods by which a sufficiently high voltage can indeed be obtained. By either of these means a higher voltage is derived from the mains voltage, on which voltage normal "TL" lamps can ignite instantaneously while the electrodes are still cold. For a lamp say 1.20 m in length and 38 mm in diameter a voltage of 400 V (r.m.s.) is required for this, whilst for longer and thinner lamps it may be as high as 750 V.

This ignition voltage is obtained by stepping up the mains voltage and when the lamp is burning a large part of it comes to lie across the ballast,

which must thus be fairly large and expensive, whilst the power factor becomes unsatisfactory.

By the second method for getting a higher ignition voltage from the mains voltage a resonance circuit is employed, but owing to the rather heavy cost of the ballast very little use has been made of this method so far.

Fluorescent lamp with low ignition voltage ("TL" S lamp)

If it were possible to construct "TL" lamps igniting on a voltage below the peak voltage of 220 V mains then the starter could be dispensed with and there would be no need to have recourse to either of the means mentioned above. Now it has indeed proved to be possible to make "TL" lamps (called "TL" S lamps) of the normal dimensions with a sufficiently low ignition voltage by applying — in a manner to be described later — on the inside of the tube a conducting strip in the longitudinal direction and connecting it electrically to one of the electrodes (*fig. 3*). The resistance of this strip, as we shall presently see, must be of the order of a few thousand ohms.

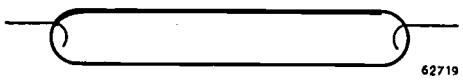


Fig. 3. Schematic cross section of the new "TL" lamp ("TL" S lamp) with conducting strip on the inside of the glass tube making contact with one of the electrodes.

This system may be understood to work in the following way. Upon the current being switched on an initial glow discharge takes place between the free end of the strip and the electrode opposite it (here the field strength is greatest). The glow-discharge current flows only during the intervals in which the electrode just mentioned is negative with respect to the strip. In the intervening intervals the discharge extends farther towards the other electrode to which the strip is connected, thereby increasing in intensity each time owing to the reduced resistance of the strip, until ultimately an arc is struck between the main electrodes.

In conformity with this explanation is the fact that in series with a choke the new starter-less lamp does not ignite so well as when it is connected in series with an incandescent lamp. Whereas the resistance of the cold incandescent lamp is negligible compared with the resistance of the strip, such is not the case with the impedance of the choke, so that when the choke is used the glow-discharge current is smaller and it is more difficult for an arc to be struck than in the case of the incandescent lamp being used.

The interval of time elapsing between the glow discharge and the normal discharge across the main electrodes depends upon the mains voltage and the resistance of the strip, as is shown by

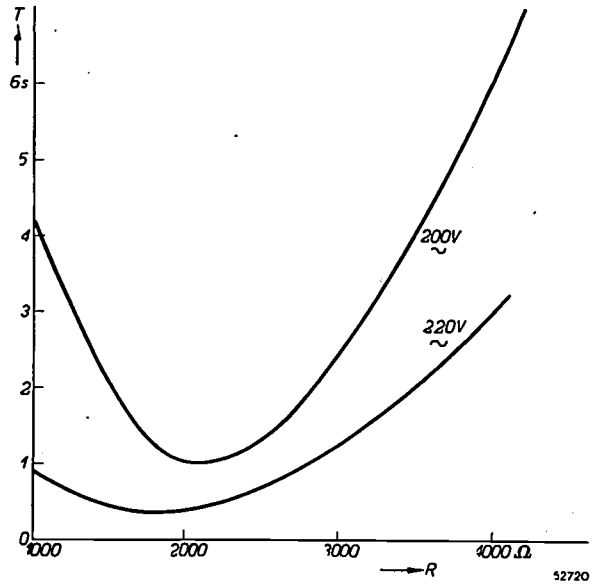


Fig. 4. The delay time T between switching on and the ignition of a "TL" S lamp of 40 W (in series with an incandescent lamp) as function of the resistance R of the strip, with 220 V and 200 V A.C. mains.

fig. 4. With resistances between 1500 and 2500 ohms and a mains voltage of 220 V the delay time varies very little with the resistance and is short, being of the order of 0.4-0.6 sec, and thus scarcely any longer than the delay when employing a rapid-action starter (see the article quoted in footnote [2])). With a resistance of 2000 ohms and a 10% too low mains voltage the delay is about 1 second. Even on 180 V A.C. voltage the lamp ignites within 10 seconds, which is usually regarded as the extreme limit. Consequently there is a wide margin for mains voltage fluctuations.

When choosing the optimum resistance one must make allowance for the loss in the strip while the lamp is burning. During the intervals in which the strip is positive with respect to the electrode opposite its free end this acts as a positive probe, with current flowing through it. In the other half-cycles it attracts positive ions towards it from the discharge, but owing to the inertness of these ions this effect is much smaller than the other. The losses in the strip appear to be roughly in inverse proportion to the resistance of the strip (see *fig. 5*, curve *1*), so that it would be desirable to have the highest possible resistance, but a limit is set to this by the delay just mentioned. For the average value of the resistance in the case of the

"TL" S lamp of 40 W 2000 ohms has been chosen, at which value the loss in the strip is still acceptable (4.5 W).

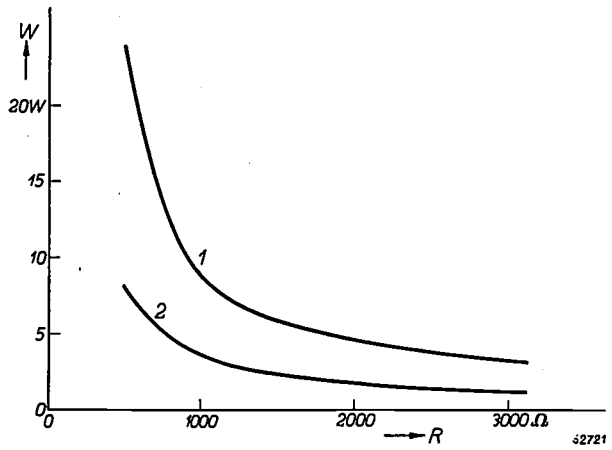


Fig. 5. The loss W in the strip of a burning "TL" S lamp of 40 W as a function of R, *1* when the strip is connected to one of the electrodes, *2* when the strip is disconnected.

When the connection between the strip and the electrode is interrupted the loss in the strip is very much less; see curve *2* in fig. 5. This is understandable, since the electron current attracted towards the strip must then at any moment be equal to the positive ion current, which, as mentioned above, is small. To take advantage of this possibility of reducing the losses one could introduce in the lamp a bimetal contact which in the cold state connects the strip to one of the electrodes (this being necessary for the ignition, as we have seen) but breaks that connection as soon as sufficient heat is developed in the burning lamp. A minor drawback to such a system is that when a lamp is switched on again just after it has been switched off it cannot reignite immediately, because it takes some time for the bimetal contact to cool down and restore the contact.

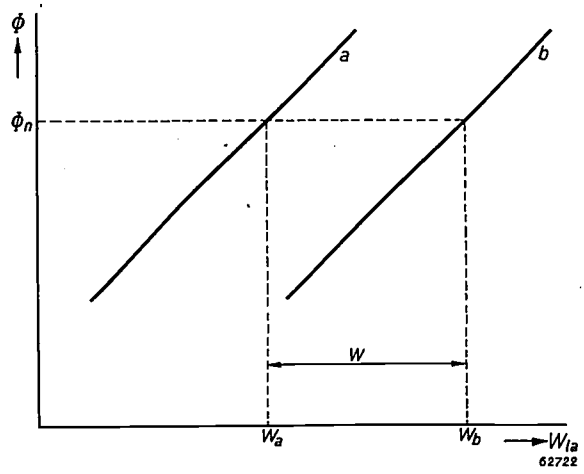


Fig. 6. Determining the loss W in the strip. The luminous flux Φ of different test lamps is measured as a function of the power consumption W_{la} in the lamp. With the exception of the strip the lamps were as nearly as possible identical. The line a gives $\Phi = f(W_{la})$ for a lamp without strip, b for a lamp with a certain strip. With $W_{la} = W_a$ and W_b respectively they yield the normal luminous flux Φ_n. $W = W_b - W_a$ is the loss in the strip.

It remains to be mentioned how the loss in the strip was measured. To this end we determined the luminous flux of test lamps with and without strip as a function of the power consumed. The strips in these lamps had different resistances and were provided with a separate lead-in so that measurements could also be taken with the strip disconnected. As a result we found almost parallel straight lines, from which the loss figures are easily derived (see *fig. 6*).

In coal-mines, where explosive gases may occur, "TL" lamps are not in themselves dangerous, but it appears that some risk attaches to the starters [3]). The "TL" S lamp, working without starter, eliminates any such possible cause of explosion.

The electrodes

The electrodes of the "TL" S lamps have to withstand heavier loading when the lamp is switched on than those of a normal "TL" lamp, since the latter lamp does not ignite until the electrodes are heated and thus ready for emission, whereas the "TL" S lamp ignites while the electrodes are still almost cold. Moreover the glow discharge between switching on and arcing causes the cathode to be bombarded with high-velocity ions.

The "TL" S lamps has therefore been fitted with special electrodes better able to withstand such conditions as these. They consist of a coiled electrode containing a larger amount of emitting material than that usually found in the electrodes of normal "TL" lamps, while the electrodes are surrounded either by a metal ring or by a closely fitting cylinder which intercepts the ions on the outside and any atomized or evaporated emitting material on the inside. Thanks to these measures the "TL" S lamp lasts on an average as long as a normal "TL" lamp.

Use on direct current

D. C. mains of 220 V are still to be found in some parts and when "TL" lamps are used on these mains they must of course be provided with a series resistance, for instance in the form of an incandescent lamp. A D.C. starter is also required,

[3]) G. D. Rieck, The illumination of coal-mines and the attendant risk of explosions, Philips Techn. Rev. **10**, 334-337, 1948/49.

in addition to a choke (the latter only for supplying a voltage surge when the starter contact opens). But there is yet another complication. While the lamp is burning the positive ions — i.e. the mercury ions — travel, on the average, towards the cathode, and thus with the constant direction of current the mercury is gradually displaced towards the surroundings of the cathode. After the lamp has been burning about ten hours there is so little mercury vapour left in the anode half that hardly any light at all comes from that part of the lamp (it is to this mercury vapour that the ultra-violet rays, which are converted into light by the fluorescent lining, are to be attributed). The discharge in that part of the lamp is then carried by argon but this gas produces only few rays to which the fluorescent phosphors are susceptible. For this reason the mains switch is often made in such a way that every time the lamp is switched on the polarity is reversed.

"TL" S lamps have proved to ignite well on 220 V D.C. mains provided the strip is connected to the positive pole of the mains. There is no need for any starter and choke as required with an ordinary "TL" lamp, but the necessity of reversing the polarity now and again remains. For the lamp to ignite on either pole it must have two strips, one connected to each electrode.

Such a D.C. lamp with two strips can be used also on A.C. mains, but then the losses in the strips are twice as great as those of an A.C. lamp with one strip of the same resistance.

Manufacture of the "TL" S lamp

The strip consists of a mixture of graphite and an enamel applied on the inside of the tube in the following way. The glass tube from which the lamp is to be made and which has already been lined with the fluorescent layer is held in a horizontal position and a small wheel is passed through it. The lower part of this wheel runs through a viscous mixture of graphite, enamel and a binder, and as it rolls along inside the tube its upper part paints the strip along the wall. The composition of the mixture and the thickness of the strip applied to the tube are so chosen that after further processing the 120 cm long and 2 mm wide strip has the aforementioned resistance of about 2000 ohms. The tube then passes through a sintering process, this being necessary for the preparation of the fluorescent layer, and in that process the binder is released from the strip, while the enamel melts and the strip is baked onto the wall of the tube.

A pinch with an electrode mounted on it is then sealed in at one end of the tube. One pole of the electrode [4]) carries a chrome-iron leaf spring (with expansion coefficient equal to that of the glass) which makes contact with the strip when, in the sealing process, the glass of the tube falls inward a little, assuming the shape indicated in *fig. 7*. At the same time the enamel of the strip is softened, so that when the tube cools down the spring is firmly secured in the enamel. Instead of the chrome-iron spring a bimetallic relay can be built in which

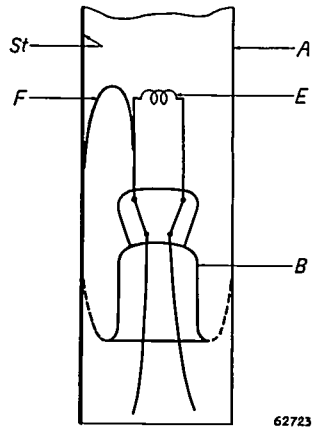


Fig. 7. *A* glass tube with fluorescent lining and conducting strip (*St*) on the inside. *B* mount with electrode *E* and chrome-iron contact spring *F*.

makes contact with the strip in the cold state but not in the hot state, as already mentioned, or else it can be introduced between the spring and the electrode.

At the other end of the tube a mount is fused in without a contact spring (unless it is a D.C. lamp, in which case it must have two strips). The lamp is then further finished off in the normal way.

Safety caps and holders

Since, contrary to the case with an ordinary "TL" lamp, current can pass through the "TL" S lamp on 220 V without the intermediary of a starter, special measures have to be taken to avoid accidents when mounting the lamp in its holders. If one end of the lamp were to be placed in the holder while the mains switch is in the "on" position and the contact at the other end of the lamp should be touched with the hand, while the fitter himself is more or less earthed, the current (the glow-discharge current of the strip) passing through the man's

[4]) Although for the normal working of a "TL" S lamp — in contrast to an ordinary "TL" lamp — only one connection per electrode is needed, both sides of the electrodes are supplied with lead-in wires so as to be able to heat the electrodes in the further manufacturing process for degassing and further processing the emitter.

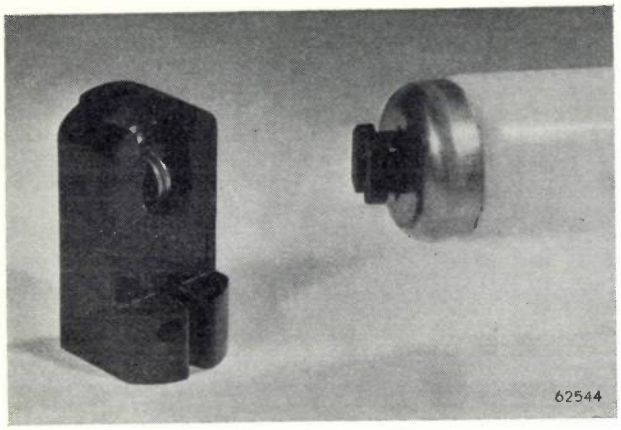


Fig. 8. Right: "TL" S lamp with safety lamp cap. *Left*: the corresponding lamp holder.

body might be dangerous. It is true that such a situation will seldom arise, because for the lamp to be inserted properly — assuming it has the normal "TL" caps — both caps have to be placed in their holders and the lamp then turned about its axis before the lamp contacts can become alive. Nevertheless with these normal caps it is still possible to turn the lamp while one cap is in the holder and to touch the live contact pins at the other end.

Means have therefore been sought to avoid the possibility of such accidents occurring. *Fig. 8* shows a new construction of lamp cap and holder. The insulating cap has a slot in which the contact (the "TL" S lamp needs only one contact at each end) is sunk so deep that it cannot be reached with what is known as a standard finger (a metal rod of standardized dimensions, in the form of a finger, used to see whether any live point can be touched with a finger). Furthermore the holder is so constructed that the circuit cannot be closed before the lamp has been turned; thus without a lamp in the holders the lamp-holder contacts cannot become alive.

Thanks to these precautions the "TL" S lamp can therefore be placed in its holders with perfect safety.

The fixture

Fig. 9 shows a fixture carrying two "TL" S lamps of 40 W and the two auxiliary incandescent lamps [5]). On 220 V A.C. mains these four lamps together yield a luminous flux of 5600 lumens for a consumption of 215 W. The photograph in *fig. 10* shows how these fixtures are used in practice.

Fig. 10. Room in Philips' head offices at Eindhoven illuminated with lamps in fixtures of the type XA 240 (fig. 9).

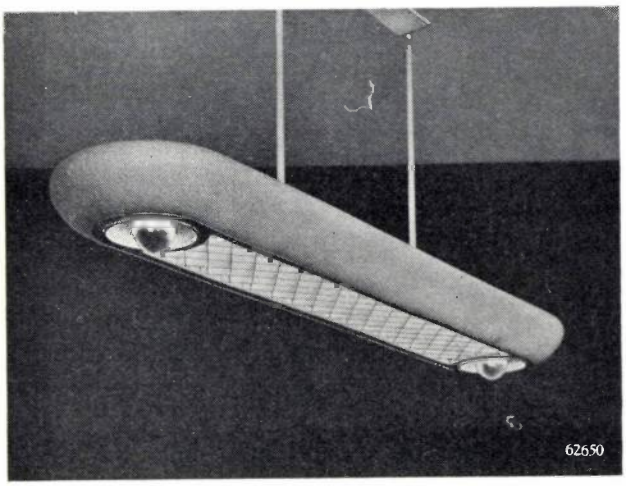


Fig. 9. Fixture (type XA 240) of "Plastocel" with two "TL" S lamps of 40 W and two incandescent lamps.

The fixture consists of "Plastocel" (a highly diffusing and washable fibrous material) and a metal frame, which, owing to the absence of any heavy parts like chokes, can be kept very light in weight. The luminous flux from the two incandescent lamps is directed downward by reflectors, so that the illumination obtained on a working plane underneath the fixture is greater than that obtained from an ordinary "TL" fixture supplying the same

[5]) For A.C. mains of 220 V an incandescent lamp for 135 V, 0.5 A, is needed, and for 220 V D.C. mains one of 110 V, 0.36 A. Since these incandescent lamps are exposed to relatively greater voltage fluctations than when connected to the mains direct, lamps of a special type must be used.

total luminous flux. The "TL" S lamps provide mainly the diffuse light. *Fig. 11* gives the distribution of the illumination on a working plane 2.5 m below the fixture. This distribution has proved to be very satisfactory, for instance, for shop lighting.

Finally it is to be noted that owing to the single-pole end contacts of these "TL" S lamps and the absence of starters the wiring in the fixture is extremely simple.

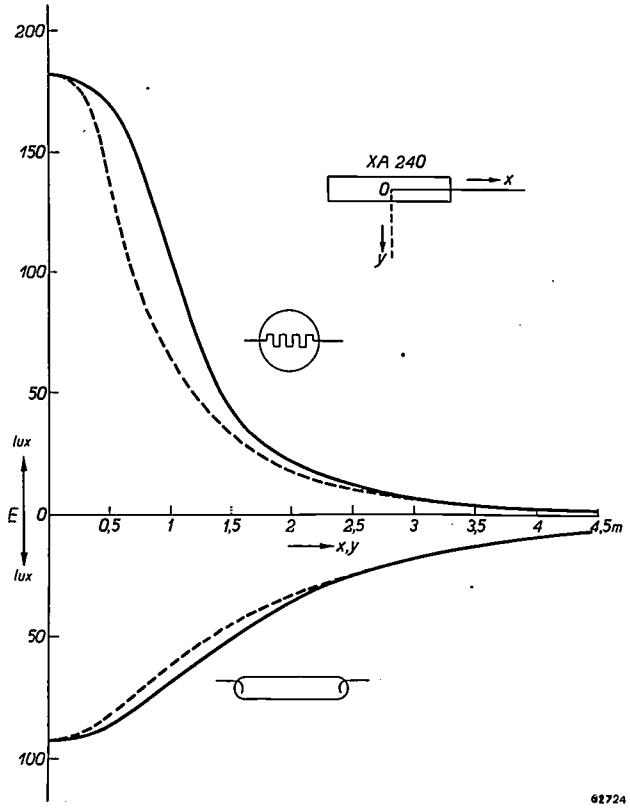


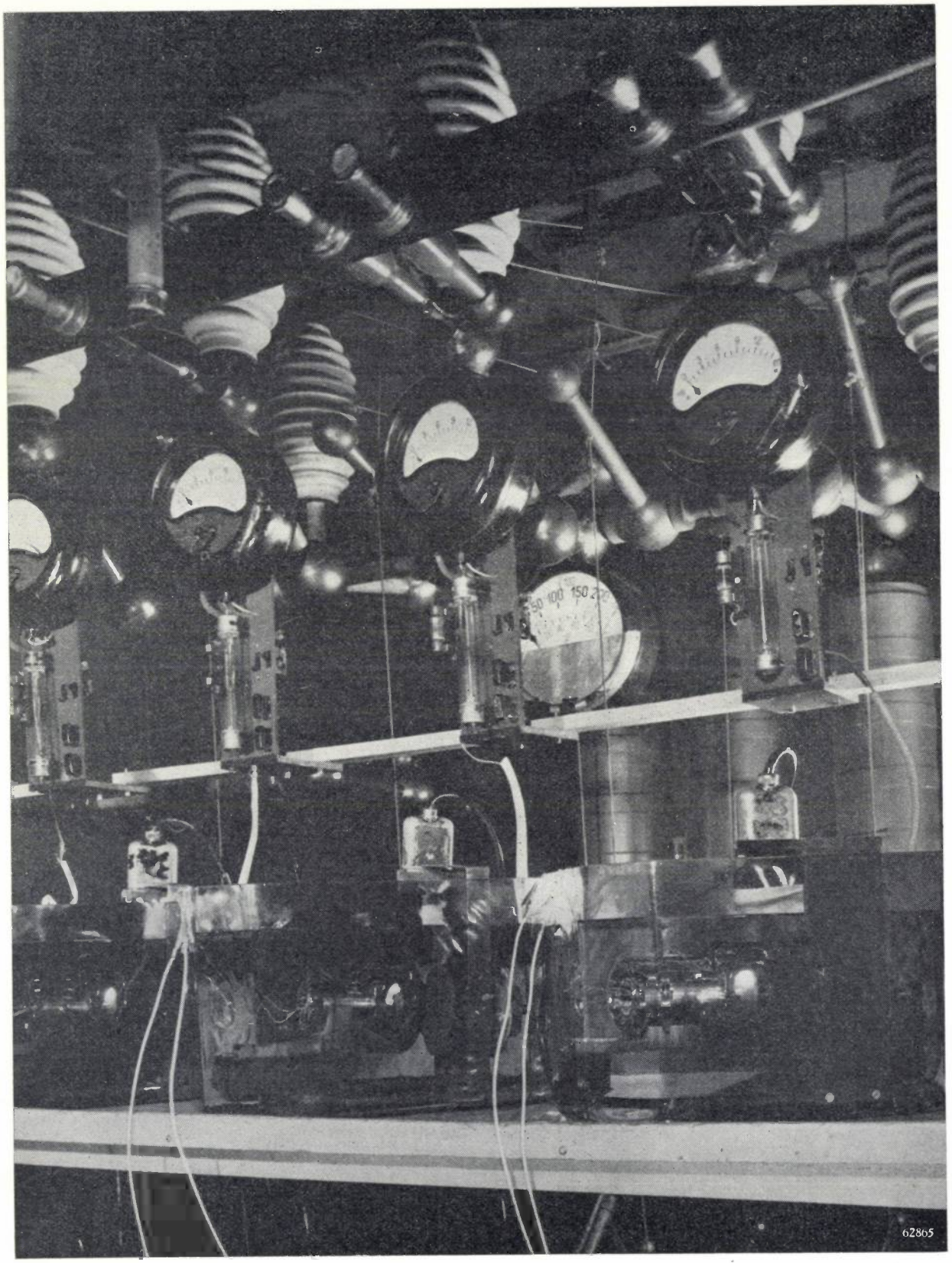
Fig. 11. Illumination E (with a luminous flux of 5600 lm) on a working plane 2.5 m below an XA 240 fixture, as a function of the distance x, y from the centre parallel to the longitudinal direction (full line) and at right angles thereto (dotted line). The contribution from the incandescent lamps has been plotted upwards, that from the "TL" S lamps downwards; the distance between the curves is thus a measure for the total illumination. Immediately underneath the fixture there is 280 lux, i.e. 185 lux from the incandescent lamps and 95 lux from the "TL" S lamps. Already at a fairly short distance from the centre the contribution from the "TL" S lamps is greater than that from the incandescent lamps.

Flickering

The asymmetry in the A.C. design of the "TL" S lamp — due to one of the electrodes being connected to a conducting strip — results in the average luminous flux in one half-cycle differing somewhat from that in the other half-cycle. Consequently the fundamental frequency of the flickering common to all lamps operating on A.C. is not twice the mains frequency but the mains frequency itself. Although this fluctuation of the light with the mains frequency is only small in amplitude the effect is still noticeable because the eye is more sensitive to flickering at a frequency of 50 c/s than to flickering of 100 c/s. When, however, two "TL" S lamps are used in one fixture, as in fig. 9, they can be so inserted that the strips are connected to different poles of the mains (it can be seen from the lamps at which end the strip is connected), the two flickerings with the mains frequency then compensating each other, so that the result is the normal imperceptible flickering with twice the mains frequency.

Summary. In some cases it is advantageous to replace the ballast of tubular fluorescent lamps ("TL" lamps) for 220 V A.C. by a special incandescent lamp, because, for one reason, the latter is cheaper and lighter in weight. If this is done with ordinary "TL" lamps, using the customary circuit with starter, there is the difficulty of the incandescent lamp being strongly overloaded each time the current is switched on, whilst moreover the ignition is unreliable. These difficulties have been overcome by the construction of a new lamp (the "TL" S lamp) provided with a strip, connected to one of the electrodes, along the inside of the glass tube (resistance of the strip about 2000 ohms). The ignition voltage of this lamp, with cold electrodes, is so low that it can be relied upon to ignite even at 180 V A.C. without any voltage surge. No starter is therefore needed. On 220 V the lamp ignites within 0.4 - 0.6 sec after switching on. For D.C. mains of 220 V a special construction of this lamp is required, with two strips each connected to an electrode. — The lamp caps and holders are so constructed that no live parts can be touched when inserting the lamp. A fixture of "Plastocel" is described for two "TL" S lamps of 40 W and two incandescent lamps, the four lamps together yielding 5600 lumens for a consumption of 215 W.

TESTING "ROTALIX" X-RAY TUBES



In this installation four X-ray tubes are tested at a time, being subjected to loads of different voltages, currents and times. The tubes, three of which are seen in the photograph, are placed in the oil-filled glass containers seen at the bottom (imitating actual working conditions). Tube currents are measured by the four meters, underneath each of which is a tubular glow-lamp serving as an indicator for any electrical disturbance in the X-ray tube. The high tension, the filament voltage and the filament current are measured by separate meters (not visible here).

THE COLOUR TRIANGLE

by W. de GROOT and A. A. KRUITHOF.

535.617

The coloured plate appended to this article is a reproduction of a picture in oils representing the chromaticity diagram in ICI coordinates. By way of explanation a brief account is given of the formation of the colour space and of the colour plane.

Conception of luminance (objective brightness) and colour

A luminous surface emitting per unit of area, per unit of solid angle and per unit of time an energy $E(\lambda)\Delta\lambda$ in the wavelength range $\lambda \rightarrow \lambda + \Delta\lambda$ in the direction of the eye has an (objective) brightness or luminance:

$$L = K\int E(\lambda)V(\lambda)\,\mathrm{d}\lambda. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The integration has to be extended over the whole of the spectrum of the electromagnetic radiation ($0 < \lambda < \infty$). K is the photometric radiation equivalent (about 650 lm/W) and $V(\lambda)$ the visibility function. The function $V(\lambda)$, which assumes values differing from zero only in the interval $400\ \mathrm{m}\mu < \lambda < 780\ \mathrm{m}\mu$, is an average derived from photometric tests with a large number of observers and standardized by the International Commission on Illumination (I.C.I.) in 1924.

Two luminous surfaces with spectral distributions $E_1(\lambda)$ and $E_2(\lambda)$ are said to have equal luminance when

$$\int E_1(\lambda)\ V(\lambda)\,\mathrm{d}\lambda = \int E_2(\lambda)\ V(\lambda)\,\mathrm{d}\lambda. \quad . \quad (2)$$

For an observer whose individual visibility function $V_{\mathrm{ind}}\ (\lambda)$ corresponds to $V(\lambda)$ these surfaces will in fact appear to be of equal brightness.

Formula (1), from which L can be calculated if $E(\lambda)$ is known, also enables us to compare the luminances of lighted surfaces objectively. All that is needed is a photocell fitted with such a filter that the sensitivity for radiations of different wavelengths depends in the same way upon λ as $V(\lambda)$; the current flowing through the photocell when it is exposed to the radiation from the luminous surface is then a measure for the quantity $\int E(\lambda)\ V(\lambda)\,\mathrm{d}\lambda$ and thus also for the luminance L.

Apart from calculation and the described measurement with the aid of a photocell, luminances can also be compared subjectively according to known photometric methods, at least provided an observer is available whose individual visibility function corresponds closely to the standardized visibility function $V(\lambda)$. One then makes use, in fact, of formula (2), where for instance E_1 is proportional to the spectral energy of the luminous surface to be measured and E_2 refers to a perfectly diffusing surface, illuminated with the standard source of light, which is compared with the first surface.

There is a certain anology between the mathematical description of the conception of "luminance" (eq. (1)) and that of the conception of "colour".

It has in fact been proved that the colour of a luminous surface can be fully characterized by three numbers denoted by the letters X, Y and Z, whereby

$$\left.\begin{aligned} X &= \int E(\lambda)\ \overline{X}(\lambda)\ \mathrm{d}\lambda, \\ Y &= \int E(\lambda)\ \overline{Y}(\lambda)\ \mathrm{d}\lambda, \\ Z &= \int E(\lambda)\ \overline{Z}(\lambda)\ \mathrm{d}\lambda. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (3)$$

These integrals bear the same character as the integral (1), $\overline{Y}(\lambda)$ even being identical with $V(\lambda)$, so that, with a factor, Y represents at the same time the luminance L. The functions $\overline{X}(\lambda)$, $\overline{Y}(\lambda)$ and $\overline{Z}(\lambda)$ (called the distribution curves or tristimulus values for the "spectrum of constant energy") were standardized in 1931 by the I.C.I. as a result of extensive tests with a number of observers. For every wavelength they have a positive value and are graphically represented in *fig. 1* [1]).

Two luminous surfaces *1* and *2* are said to have the same colour (i.e. are indistinguishable) when

$$X_1 = X_2,\ Y_1 = Y_2 \text{ and } Z_1 = Z_2\,.$$

[1]) In accordance with P. J. Bouma, for these functions we have introduced the notations $\overline{X}$, $\overline{Y}$, $\overline{Z}$ (capitals) because they appear to be more logical than the ususl notation $\bar{x}$, $\bar{y}$, $\bar{z}$ (in small letters).

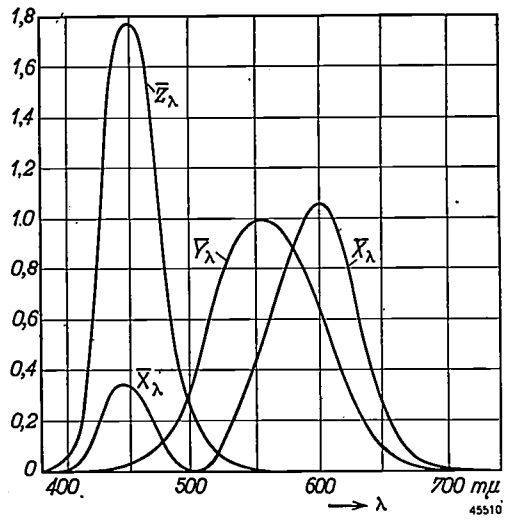


Fig. 1. The functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$, the so-called spectral distribution curves or tristimulus values (for the spectrum of constant energy).

We assume the theory of Young and Helmholtz to be correct, according to which colours are perceived by means of three kinds of receptors present in the retina, having hypothetical spectral sensitivities $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$. Then the functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ must be linear combinations of $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$. There is still, however, some uncertainty about the coefficients in these linear relations and thus about the functions $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$.

The mutual deviations of the functions corresponding to $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ of different observers are relatively small, at least if certain groups of persons — called protanopes and deuteranopes — showing decidedly systematic aberrations are excluded; such people are not suitable for carrying out subjective colorimetric measurements which should hold for normal persons.

The fact that the individual visibility function $V_{\text{ind}}(\lambda)$ of different, colorimetrically normal, observers differs from the standardized average function $V(\lambda)$ much more than the individual function corresponding to $\bar{y}(\lambda)$, does not detract from the truth of the statement that $V(\lambda)$ and $\bar{y}(\lambda)$ are identical, since $V_{\text{ind}}(\lambda)$ may be regarded as being built up from the functions $R(\lambda)$, $G(\lambda)$ and $B(\lambda)$ and thus also from the functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$. Thus

$$V_{\text{ind}}(\lambda) = a\bar{x}(\lambda) + b\bar{y}(\lambda) + c\bar{z}(\lambda),$$

in which the standardized functions may be taken for $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ and the coefficients a, b and c differ as between one person and another [2]). On the average $a = c = 0$ and $b = 1$.

The numbers X, Y and Z characterizing a colour can be determined, just like L, in different ways.

In the first place, when $E(\lambda)$ has been measured, X, Y and Z can be calculated from (3).

In the second place X, Y and Z can also be determined by objective measuring, namely by using three photocells each with such a filter that the sensitivities of the combinations are proportional to $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ respectively. Such objective colorimetric measuring methods are in fact employed.

Finally, X, Y and Z can also be determined subjectively, by making in a suitable instrument (colorimeter) the colour of the surface concerned indistinguishable from that of another, white, reference surface illuminated with an additive mixture of three coloured standard sources of light each of which gives that surface respectively the luminance L_1, L_2 and L_3. X, Y and Z can then easily be calculated from the three luminances L_1, L_2 and L_3, or found, with the aid of the tables or graphs supplied with the instrument, from the settings a_1, a_2 and a_3 of the control knobs with which L_1, L_2 and L_3 were brought to the right values.

The colour space

When the values of X, Y and Z have been determined for a large number of colours one can imagine these as being plotted as coordinates in a tri-axial system of coordinates. The colours are then said to be represented in a colour space. In this colour space a special part is played by colours of which the representative points lie on an arbitrary line drawn through the origin. Such a series of colours is seen when a white source of light is brought closer and closer to a coloured surface (or inversely a coloured light is brought closer to a "white" surface). The closer the light source is brought to the surface the greater is the amount of radiation $E(\lambda)$ reflected from that surface, the ratios $X : Y : Z$ remaining unchanged. In plain words one would say that as the lamp is brought closer the "colour" does not change but becomes brighter. We prefer, however, to say that two colours differing only in brightness are "different". For what does not change in the test described we introduce the name colour species or chromaticity. We can then describe the said test with the words: The colours perceived when a lamp is brought closer to a surface differ in brightness but belong to the same colour species (have the same chromaticity).

The colour plane

Since a colour is fully determined by three quantities, it is obvious that two quantities are sufficient for determining a chromaticity. For this purpose two of the three quantities x, y and z are taken, which are given by:

[2]) H. de Vries, Physica, The Hague 14, 319-366, 367-380, 1948.

$$\left.\begin{aligned} x &= \frac{X}{X+Y+Z}, \\ y &= \frac{Y}{X+Y+Z}, \\ z &= \frac{Z}{X+Y+Z}, \end{aligned}\right\} \quad \ldots \quad (4)$$

and hence

$$x + y + z = 1.$$

From the fact that two quantities are sufficient it follows that chromaticities can be represented in a plane. If, for instance, x and y are plotted as coordinates in a rectangular system of axes then all colour points [3]) are contained within the triangle determined by the x-axis ($y = 0$), the y-axis ($x = 0$) and the straight line $x + y = 1$; in fact, since $\overline{X}(\lambda)$, $\overline{Y}(\lambda)$ and $\overline{Z}(\lambda)$ are positive for any value of λ, and thus also the quantities X, Y and Z are positive, none of the three quantities x, y and z can be negative.

All kinds of chromaticities can now be localized in the plane. Inter alia, the colour point can be indicated for any monochromatic radiation with wavelength λ, since in this case $X : Y : Z = \overline{X}(\lambda) : \overline{Y}(\lambda) : \overline{Z}(\lambda)$, from which the corresponding x and y can be calculated for any wavelength. A curve drawn through the points thus obtained is called the spectrum locus. This curve lies indeed entirely within the aforementioned triangle and touches its sides at two places. It has the shape of a horseshoe (see fig. 2).

The mixture law

A peculiar property of the colour plane, as a direct consequence of the properties of the eye determined in formula (3), is the following. When a white surface is illuminated with two coloured lights simultaneously and thus, so to say, the colours which would result from the illumination with only one light source at a time are additively mixed, then the resulting point in the chromaticity diagram lies on the line connecting the colour points of the components. The position of the point representing the mixture on that line can be deduced from the well-known centre of gravity rule by allotting to each of the component chromaticities a "weight" equal to the quantity $(X + Y + Z)$ or Y/y (cf. formula (4)).

From this it follows that the colour points of all realizable chromaticities lie within the figure bounded by the spectrum locus and the straight line connecting the extremities of that curve. On this straight line, called the purple line, lie the colour points of the chromaticities obtained when the extreme spectral red is additively mixed with the extreme spectral violet.

Two colours which, if mixed in suitable proportions, give the chromaticity white (for which by definition $x = {}^1/_3, y = {}^1/_3$) are said to be complementary. In particular, any colour having its colour point on the purple line is complementary to a spectral colour ($494\ \text{m}\mu < \lambda < 570\ \text{m}\mu$).

A given colour can always be regarded as being a mixture of the aforementioned white with a spectral colour or with a colour lying on the purple line. The quotient of the required brightness of the spectral colour (or of the colour on the purple line) and the total brightness of the given colour is known as the colorometric purity p. The wavelength of the spectral colour that has to be additively mixed with white to imitate a given colour species is called the dominant wavelength. For the purple colours the wavelength of the complementary spectral colour is chosen for this purpose, this case being distinguished by a minus sign.

Realization of the colour plane

Once it is known that a point x, y in the colour plane represents a certain chromaticity one would like to know what that chromaticity looks like.

To give an idea of the distribution of the chromaticities over the colour plane, coloured strips of paper (e.g. colour samples from the Munsell atlas) could be pasted over their corresponding positions in the chromaticity diagram.

One could also determine the colour points of a number of oil paints and give a dab of these paints on the right places in the diagram. When this has been done, one could try to make the intermediate chromaticities with suitable paints and thus complete the whole of the chromaticity diagram. The attached coloured plate is a reproduction of the result of such an attempt [4]).

Of course this result is not without its defects, but when viewed from a distance the plate gives a reasonable impression of the manner in which the chromaticities are distributed over the diagram.

In one respect the plate is certainly incorrect.

[3]) For the sake of brevity the expression colour point is used also for the colour plane, although strictly we should still speak of colour-species point or chromaticity point.

[4]) We are greatly indebted to A. R. W. Muijen of this laboratory for the manner in which he has performed this task.

The fact is that if it is desired to represent highly saturated colours, paints must be used which show a reflectivity differing from zero only in a very limited spectral range and thus look very dark in comparison with less saturated colours. With the customary paints, and starting from white, one therefore does not get much farther than about halfway to the limit of the spectral curve. Only in yellow and violet is it possible to get with oil paints a rendering of chromaticities close to the edge of the range. For the sake of the total impression the plate has, however, been filled in everywhere right up to the edge, be it with colours lacking in saturation.

It is well to point out that in order to get the right impression of the chromaticity diagram the reproduction should be viewed in daylight because the calculation of the colour points of the applied paints has been carried out for daylight.

For further orientation in the diagram a line drawing of the colour triangle is given in *fig. 2* in which the values of x and y are indicated, as also the wavelengths of the spectral colours and, on the purple line, the wavelengths of the spectral colours complementary to purple colours. In addition, the lines of constant dominant wavelength and the lines of constant colorimetric purity are shown.

The number of chromaticities

When trying to arrive at a colorimetric match between two colours *1* and *2*, where, for instance, *1* is a given colour and *2* a variable additive mixture of three "primary colours", one will aim at obtaining the equalities:

$$X_1 = X_2, \quad Y_1 = Y_2, \quad Z_1 = Z_2 .$$

In practice, however, the two halves of the field of vision in the colorimeter will be indistinguishable already, while there is in fact still a difference

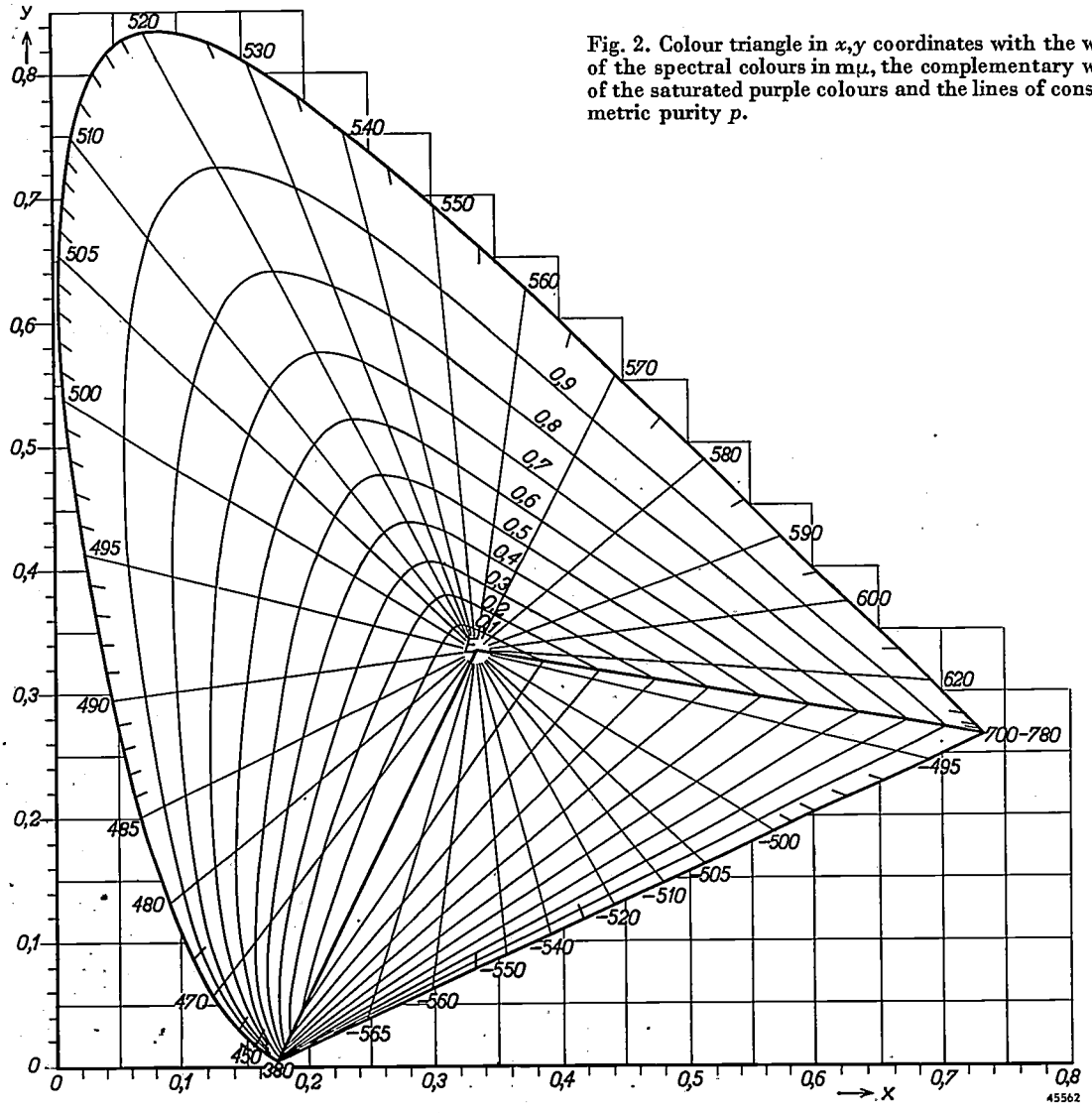


Fig. 2. Colour triangle in x,y coordinates with the wavelengths of the spectral colours in mμ, the complementary wavelengths of the saturated purple colours and the lines of constant colorimetric purity p.

(expressed by ΔX, ΔY and ΔZ) between them. Starting from a certain point x, y in the chromaticity diagram, for a given luminance and for a certain direction through that point, it can be deduced from a large number of such observations what average departure from the given chromaticity is possible along that line without disturbing the match of the two colours to a perceptible degree. Thus along each line through x, y two points are found, at either side of the point x, y. Repeating this construction for a number of directions through the same point x, y and connecting the points thereby found, a curve is obtained around the starting point. Such a curve is found to have practically the shape of an ellipse. The size and orientation of this ellipse differ for different points of the chromaticity diagram. A number of such ellipses have been determined by Mc Adam[5]). The area of the ellipse, multiplied by a suitable numerical factor, can be taken as a measure of the space occupied by a chromaticity in the chromaticity diagram. From this it is possible to calculate the number of different chromaticities, of which there appear to be over 10,000.

Colour stimulus and colour sensation

We have seen that a colour is fully determined by three data, e.g. by the coordinates X, Y and Z in the colour space or by luminance and chromaticity (determined by x and y), or by the luminance, the dominant wavelength and the colometric purity.

The question whether a colour determined in this way has the same appearance under all circumstances has to be answered in the negative. A piece of red paper, for instance, reflecting 25% of the incident light will appear to be browner against a white background than against a black one. A piece of grey paper, for which $x = {}^1/_3$, $y = {}^1/_3$, appears to be bluish against a yellow background and yellowish against a blue background, greenish against a red and reddish against a green background.

This difference between the conceptions of colour stimulus and colour sensation has been discussed elsewhere in this journal. In respect to this and other more detailed questions on the subject reference is made to the literature quoted at the end of this article.

[5]) If $A(x, y)$ is the area of the ellipse described above around the point x, y then the number of chromaticities is

$$\int \frac{dx\,dy}{k\,A(x,y)},$$

in which k is a numerical factor. The integral has to be extended over the area of realizable chromaticities.
D. L. Mc Adam, Visual sensitivities to color differences, J. Opt. Soc. Amer. **32**, 247-274, 1942 and D. L. Mc Adam, Documenta Ophtalmologica **3**, 214-239, 1949.

Producing a colour plane by means of coloured lights

The colour plane, or at least part of it, can be produced in another way by making use of the so-called first law of Grassmann, according to which any chromaticity can be perfectly imitatedb ya na dditive mixture of three arbitrarily chosen chromaticities, say a red (r), a green (g) and a blue (b), with spectral distributions $E_r(\lambda)$, $E_g(\lambda)$ and $E_b(\lambda)$, respectively. From equation (3) it follows that for this purpose three constants, K_1, K_2, K_3, have to be determined, such that

$$\left.\begin{aligned}\int E\bar{X}\,d\lambda &= K_1\int E_r\bar{X}\,d\lambda + K_2\int E_g\bar{X}\,d\lambda + K_3\int E_b\bar{X}\,d\lambda\\ \int E\bar{Y}\,d\lambda &= K_1\int E_r\bar{Y}\,d\lambda + K_2\int E_g\bar{Y}\,d\lambda + K_3\int E_b\bar{Y}\,d\lambda\\ \int E\bar{Z}\,d\lambda &= K_1\int E_r\bar{Z}\,d\lambda + K_2\int E_g\bar{Z}\,d\lambda + K_3\int E_b\bar{Z}\,d\lambda\end{aligned}\right\} \quad (5)$$

These equations can always be solved provided the coefficient determinant $\Delta \neq 0$, a condition which is complied with in the case of the colours red, green and blue. K_1, K_2 and K_3 are positive if the colour point of the colour to be imitated lies inside the triangle rgb; K_1, K_2 and K_3 may be regarded as colour coordinates. The corresponding colour space is to be regarded as an *affine* transformation of the XYZ space. In a manner similar to that described above, from this three-dimensional space it is possible to derive a two-dimensional representation of all chromaticities which in the xy-plane fall inside the triangle rgb, this being done by introducing the quantities

$$\left.\begin{aligned}k_1 &= \frac{K_1}{K_1+K_2+K_3},\\ k_2 &= \frac{K_2}{K_1+K_2+K_3},\\ k_3 &= \frac{K_3}{K_1+K_2+K_3}.\end{aligned}\right\} \quad \ldots\ldots (6)$$

As previously done for x and y, k_1 and k_2 for instance can be regarded as cartesian coordinates, or k_1, k_2 and k_3 can be plotted as triangular coordinates in an equilateral triangle (*fig. 3*). The relation between

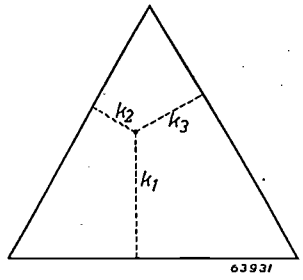


Fig. 3. Triangular coordinates k_1, k_2, k_3.

this $k_1k_2k_3$-plane and the xy-plane is given by a *projective* transformation.

In *fig. 4* a device is illustrated whereby the colours in the $k_1k_2k_3$-plane automatically occur in the right places. S_1 is a metal screen with an aperture in the shape of an equilateral triangle and covered

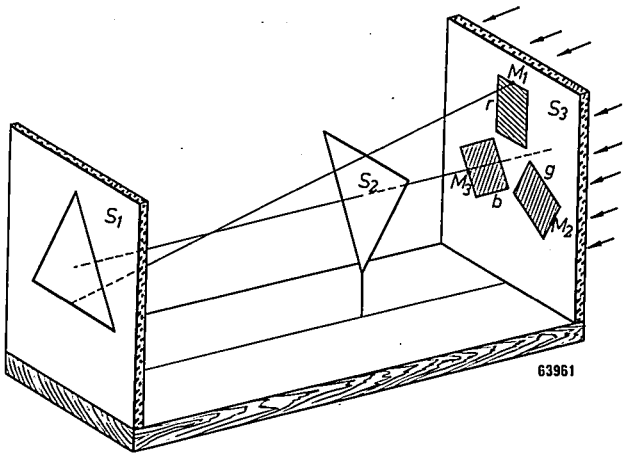


Fig. 4. Set-up for displaying the colour plane in triangular coordinates.

at the back with a plate of frosted glass. S_2 is a metal plate of the same shape as the aperture in S_1 but placed upside down. The screen S_3 has three rectangular apertures so arranged that their long sides make angles of 120° with each other. The centres M_1, M_2, M_3 of the short sides facing outward form an equilateral triangle of the same size and in the same position as the aperture in S_1. The distance S_1S_2 is twice the distance S_2S_3. The centres of all triangles lie on one line normal to the three screens. Otherwise the dimensions are arbitrary. In an actual model, for instance, the height of the triangles was 10 cm and S_1S_2 was about 25 cm. The openings r, g, b in S_3 have been covered respectively with a red, a green and a blue filter (Wratten filters 29 F, 61,47). Behind the screen S_3 is a plate of frosted glass, illuminated uniformly with a strong incandescent lamp.

Let us now ascertain what the colour is at a point P lying inside the diaphragm in the plane of S_1, having the triangular coordinates k_1, k_2, k_3 (taking the height of the triangle as equal to 1). Imagining the screen S_2 as being projected from P onto S_3, then we get on S_3 a triangular "shadow" of S_2 (*fig. 5*). Only those parts of the rectangular apertures falling outside this shadow can transmit light to P. The heights of these rectangular pieces are respectively $k_1/2$, $k_2/2$ and $k_3/2$. If the other sides are denoted by $2b_1$, $2b_2$ and $2b_3$ then these areas are equal to k_1b_1, k_2b_2 and k_3b_3. Now if $E_0(\lambda)$ is the spectral distribution of the lamp illuminating S_3 and $r(\lambda)$, $g(\lambda)$, $b(\lambda)$ represent the spectral transmission factors of the three filters, then — disregarding the effect of the obliquity of some of the rays — the illumination at P is proportional to

$$\int E_0(\lambda)\,[k_1b_1r(\lambda) + k_2b_2g(\lambda) + k_3b_3b(\lambda)]\,\mathrm{d}\lambda,$$

and thus in P we have X, Y, Z according to (5), where

$$\left.\begin{aligned} E_r(\lambda) &= b_1E_0(\lambda)\ r(\lambda) \\ E_g(\lambda) &= b_2E_0(\lambda)\ g(\lambda) \\ E_b(\lambda) &= b_3E_0(\lambda)\ b(\lambda) \end{aligned}\right\} \quad \ldots\ . \quad (7)$$

Hence k_1, k_2 and k_3 in P are proportional to $K_1K_2K_3$ and thus are identical with the quantities k_1, k_2, and k_3 as defined by (6).

The part of the xy-plane (inside the triangle rgb, see *fig. 6*) is therefore truly represented on the screen, be it projectively transformed, with r, g and b corresponding to the light transmitted by the filters. The colour at P still depends upon the quantities b_1, b_2, b_3. By varying the gap widths the colour plane can thus be made to transform within itself while retaining the colours of the angular points. By means of this variation it is still possible to make the colour plane satisfy certain conditions, for instance that a prescribed point should have the colour white ($x = 1/3$, $y = 1/3$).

When one comes to carry out the experiment it will be found that this white gives the impression of grey. The colour at a certain point of the frosted glass can best be judged by holding in front of S_1 a screen with a small aperture. It then appears that part of the colour plane does indeed give a "white" impression. The widths b_1, b_2, b_3, if chosen

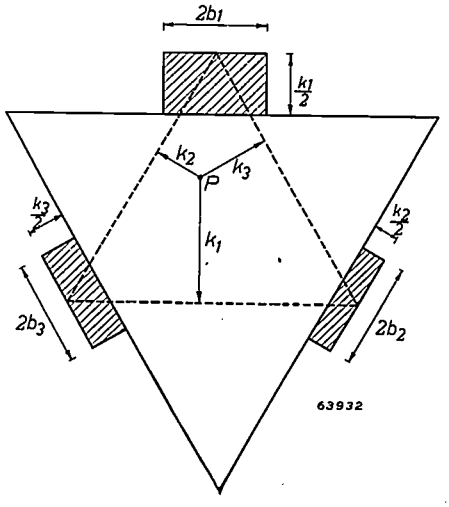


Fig. 5. How the colour observed at P is formed. To the eye with the pupil at P (imagining the frosted glass plate to be removed) only the hatched parts of the apertures r,g,b are visible.

in the right proportions, make no difference so long as the gaps do not overlap. It is therefore possible to employ also linear sources of light, the intensities of which can be adjusted separately. In principle monochromatic light sources can be used, so that the triangle *rgb* can be made to cover almost the whole of the colour plane and highly saturated colours are also truly represented, in contrast to the representation by means of paints or printers' inks.

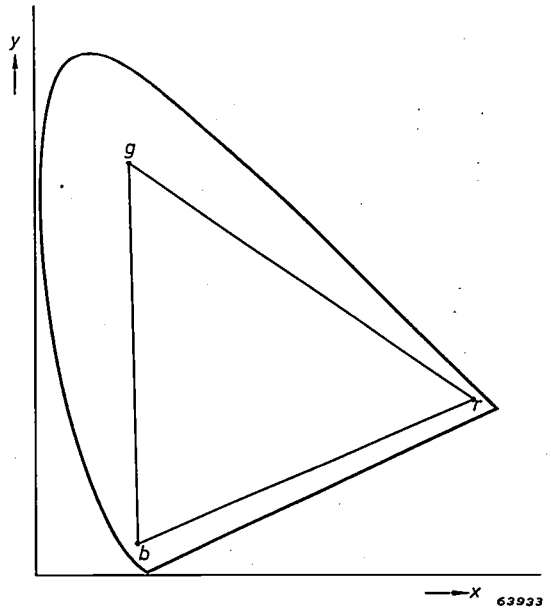


Fig. 6. The primary chromaticities E_r, E_g and E_b of formula (7) in the xy-plane.

Further, by giving the diaphragms and screens a triangular shape differing from the equilateral shape it is possible to make the colour plane congruent with a part of the xy-plane.

Bibliography

P. J. Bouma, The perception of colour, Philips Techn. Rev. **1**, 283-287,1936; The representation of colour sensations in a colour space-diagram or colour triangle, Philips Techn. Rev. **2**, 39-46, 1937.

P. J. Bouma and A.A. Kruithof, Colour stimulus and colour sensation, Philips Techn. Rev. **9**, 2-7, 1947.

J. L. H. Jonker and S. Gradstein, Fluorescent pigments as an artistic medium, Philips Techn. Rev. **11**,16-22, 1949.

P. J. Bouma, Physical aspects of colour, PhilipsTechnical Library 1948. See also Philips Techn. Rev. **9**, 158, 1947.

Summary. In analogy with the conception of luminance (objective brightness) the conception of colour is dealt with. A brief explanation is given of the construction of the colour space in XYZ coordinates (I.C.I. 1931) and of the corresponding chromaticity diagram, it then being discussed in how far it is possible to give a rendering of the colour plane with the aid of oil paints. A reproduction of the result of an attempt in that direction is appended. Finally the number of chromaticities is proved to exceed 10,000 and reference is made to the difference between the conceptions of "colour stimulus" and "colour sensation".

POSITIONING OF THE SOURCES OF LIGHT WHEN PHOTOGRAPHING WITH ARTIFICIAL LIGHT

by S. GRADSTEIN.

771.44.022

For the photographer who wishes to use artificial light instead of daylight for taking photographs there are two cardinal differences between these two kinds of light: the difference in spectral distribution (colour rendering) and the difference in the geometrical situation when illuminating the object. Much has already been written about the first difference and its consequences are taken into account in the manufacture of photographic emulsions and in the construction of lamps developed for photographic purposes. The geometrical differences are usually allowed for more or less by intuition, but it may be useful to study them rather more quantitatively.

Photography as practised at the present day has two aspects, that of an expression of art and that of documentation, with sometimes the one and sometimes the other being given prominence. The aspect of documentation is clearly seen, for instance, in press photography and in medical photography, whilst also amateurs' snaps as recordings of happy memories bear more the nature of a documentation than an expression of art. As a special case of the documentary aspect is to be considered photography as applied for purposes of reproduction.

Nowadays artificial light is largely used both for artistic and for documentary photography, one of the reasons being the desire to be independent of daylight with its limitations in time and quantity. In art photography another reason is that with artificial light better means are afforded for giving expression to an object, since there is more freedom in distributing the light over the object than is possible with daylight. With daylight the whole of the sky, or that part of it which is visible, acts as the source of light. Besides this diffuse illumination there is the directed illumination from the sun, practically a point source of light at an infinite distance, the direction of which can be chosen only to a limited extent. With artificial light, on the other hand, directed light can be projected onto the object from any direction desired, even upward, and also from different directions simultaneously, so that with the play of light and shadow it is possible to give the object one's own vision of it[1]).

As regards the choice and positioning of the sources of light, in art photography there is a wealth of experience to draw upon. In a certain book on portrait photography [2]), for example, no fewer than 64 different arrangements of lamps have been thoroughly analysed. There has been no lack of attempts to distil certain recipes from the mass of experience gathered, but recipes will only be followed by those whose artistic talents are not yet fully developed; the artist photographer will always follow his own bent, guided by the artist's natural gifts of taste and intuition, accompanied by an aversion to all that is conventional. Thus the use of artificial light becomes what might be called a "lighting art".

The case where photography bears a documentary character is a more simple matter. For the illumination of the object with artificial light, as well as that with daylight, some well defined requirements can then be formulated, which may be summarized as follows: the whole of the object must be uniformly illuminated, alternated by shadows where the plasticity of the object has to be shown to best advantage, though with not too hard shadows.

When working with daylight the first two of these requirements can be met by the sun, if it is acting as the "main" or "primary light", whilst the diffuse supplementary light from the sky — or from reflecting surfaces, such as walls — may provide for the necessary softening of the shadows. When using artificial light, too, one will always have a "primary light" supplemented by "secondary lights" or by a more or less diffuse general lighting, which may be obtained from the reflection of the surroundings. But with artificial light we run the risk of failing

[1]) Also the nature of the shadows can be influenced, namely the penumbra width, since there is a choice between sources of light of different shapes and dimensions (spot-lights, normal reflectors with or without diffusers, linear lamps, light boxes, etc.), whereas the penumbra width of direct sunlight is invariably 30 minutes of arc.

[2]) W. Nurnberg, Lighting for photography, Foco Press, London-New York, 1940.

Fig. 1. Photograph showing a disturbing inequality in the illumination, falsifying the uniformity of the military uniforms. From the fall of the shadows (and the two connecting leads on the right and left!) it is to be seen that two lamps were used, placed too close to the group, presumably owing to lack of space.

to comply even with the first requirement, viz. that of uniform illumination of the whole of the object. The lamp (or combination of lamps) serving as the primary light cannot be placed at an infinite distance, like the sun, with the result that there are perceptible differences in the distances between the lamp and the various parts of the object and a certain inequality arises in the illumination of those parts.

That this risk of a discordant inequality is not imaginary can be seen from *fig. 1*, a photograph taken by a serious amateur photographer. A diffuse supplementary lighting reduces this danger, since an additive contributory illumination equal for all points reduces relatively the differences in the illumination from the main light. But also in this respect one is at a disadvantage compared with the use of daylight, because for reasons of economy (the high values of luminous flux required!) one must invariably manage with a very much smaller degree of diffusion of the illumination than is to be obtained with daylight.

Roughly speaking, the inequality in the illumniation can always be sufficiently reduced if the lamps [3]) are placed far enough away from the object. On the other hand, however, — even when the fact that sometimes the space available does not permit of any great distance is disregarded — it is desired to place the lamps close to the object so as to make the most economical use of the light or to have the shortest possible exposure. It is therefore well worth while investigating what minimum distances are needed in order to satisfy the requirement of a sufficiently uniform illumination.

In this article we shall consider from this point of view some simple arrangements of lamps. The practical rules that will result from this investigation cannot, of course, substitute the experience that every photographer must himself acquire to be able to work successfully with artificial light,

[3]) Unless otherwise stated, where the word lamp or lamps is used henceforth this refers to the main light.

but they may well serve him as a guide for his experiments and they may particularly be of use in cases where the lamps are permanently fixed for one and the same kind of photographic work, like that of the professional photographer and that for reproduction purposes.

A single lamp directly facing the object

Distance requirements

Let us begin with the simple case where ouly one lamp is used and assume, first, that it is placed centrally in front of the object (of course slightly higher or lower than the camera). The lamp may be regarded as a point source of light with the same luminous intensity I in all directions. A small surface at a distance R perpendicular to the rays of light then receives the illumination

$$E = \frac{I}{R^2} \quad \ldots\ldots\ldots \quad (1)$$

Suppose that the object to be photographed has a certain depth d (*fig. 2*). The foremost parts, at a distance r from the lamp, receive the strongest illumination and those at the back the weakest;

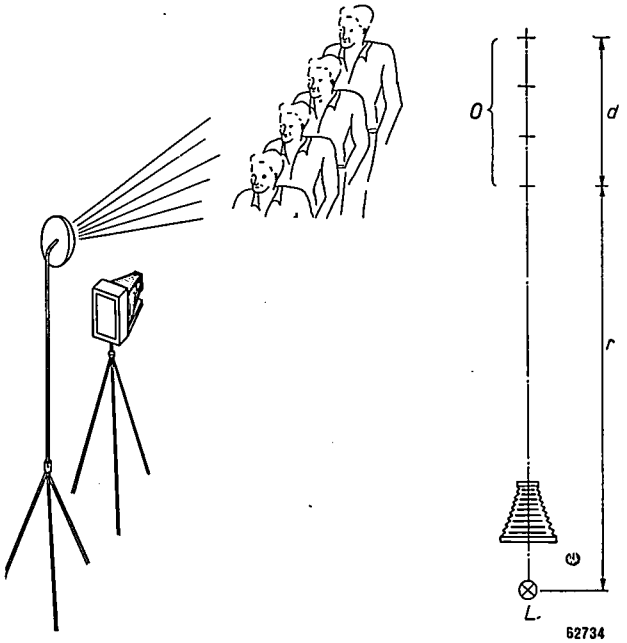


Fig. 2. A lamp L is placed at a distance r in front of the object O having a depth d.

the ratio of the two values of illumination is $K = (r + d)^2/r^2$. Allowing a certain maximum value for the "inequality factor" K, we can immediately calculate the minimum permissible value of the quotient r/d. In *fig. 3*, r/d has been plotted as a function of K.

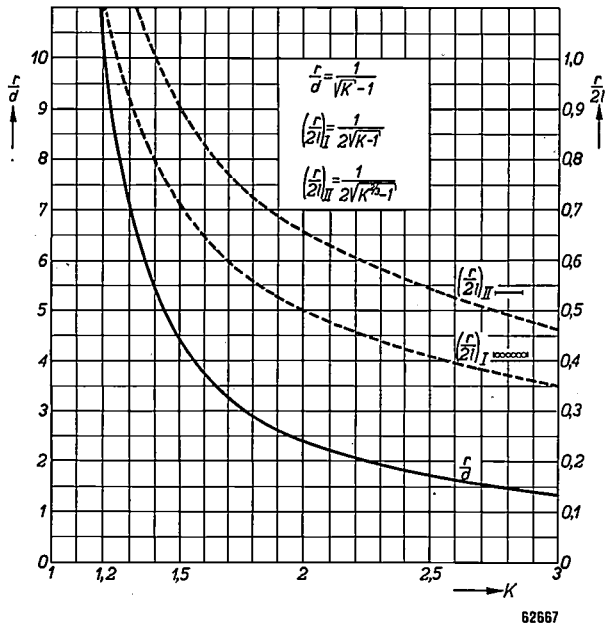


Fig. 3. Relation between the inequality factor K of the illumination and the value of r/d (fig. 2) or $r/2l$ (figs 4*a* and *b*).

It will depend upon the purpose of the photograph what value of K will in practice be considered permissible. To confine our thoughts let us take this permissible value of K as being 1.5, in which case we get:

$$r_{\min} = 4.5\, d. \quad \ldots\ldots\ldots \quad (2)$$

It is reasonable to assume that in most cases an inequality factor of $K = 2$ will be judged to be rather high. On the other hand an illumination with $K = 1.2$ may be considered to be very uniform. It is only in reproduction work that values of K less than 1.2 are required.

Next to be considered is the width (or height) of the object to be photographed. Suppose that we have an object consisting of a number of persons standing side by side in a row of the length $2l$ (*fig. 4a*). The illumination at the ends of the row is a factor $K = (r^2 + l^2)/r^2$ less than that in the middle. The minimum permissible value of $r/2l$ as a function of K has likewise been plotted in fig. 3 (curve $(r/2l)_I$). For $K = 1.5$ we find:

$$r_{\min} = 0.7 \times 2l. \quad \ldots\ldots\ldots \quad (3)$$

Here it is assumed, in this case too, that the surfaces in each part of the object are perpendicular to the rays of light. With a group of people as the object there are indeed surfaces in this position for each person, though they are not identical as between one person and another. We must, however, also reckon with objects where the principal surfaces to be illuminated are all parallel, an extreme example of which is an entirely flat object (fig. 4*b*) like

a painting or a page of a book that has to be reproduced. In such a case the outer edges of the object are at a still greater disadvantage compared with the part in the middle, on account of the fact that they are not only farther removed from the

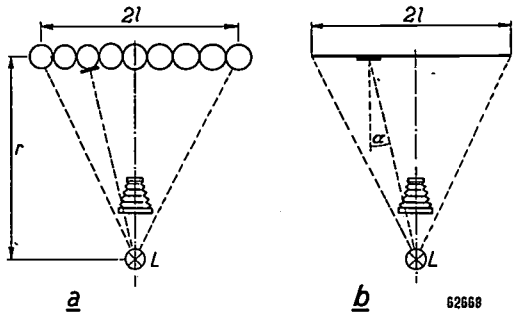


Fig. 4. A lamp L placed centrally in front of an object with a width $2l$. Where the object is a series of spheres or a row of people (a) then in each part of the object there are small surfaces perpendicular to the rays of light; with a plane object, such as in reproduction work (b), the surfaces in different parts of the object stand more or less at an angle to the rays of light.

source of the light but also stand at an angle instead of being perpendicular to the rays of light. For the illumination of a surface whose normal makes an angle α with the rays of light the formula (1) is replaced by the more general formula:

$$E = \frac{I}{R^2} \cos \alpha . \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

For the edge of the object we have $\cos \alpha = r/\sqrt{r^2+l^2}$. The ratio for the illumination in the middle and that at the edge is now:

$$K = \left[\frac{r^2 + l^2}{r^2} \right]^{3/2}.$$

For $K = 1.5$ we then find the more stringent requirement:

$$r_{\min} = 0.9 \times 2l \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Also for this case $r/2l$ has been drawn in fig. 3 as a function of K: curve $(r/2l)_{II}$.

The conditions contained in equations (2) and (4) have already been put forward by Van Liempt[4]). Some further considerations will be devoted to this part of our subject, followed by a discussion of the more general cases, viz. those of one lamp to the side facing the object and of two symmetrically placed lamps [5]).

Practical significance of the set-up considered

The method of illumination with a single lamp directly in front of the object is not commonly applied because, owing to the fall of the shadows, it does not lend any great plasticity to the photographs. This objection applies especially when photographing one single three-dimensional object, as is the case in portrait photography. Neither will this method be generally followed in the making of reproductions, on account of the disturbing reflexions.

Nevertheless there are cases where this arrangement is of importance, as in colour photography, where — with some exceptions — a plastic effect is not obtained primarily by means of shadows but rather with the aid of colour contrasts. It should be noted that in colour photography both high uniformity in the lighting and a relatively strong illumination are required, so that the foregoing is particularly applicable to this kind of photography.

Another case where the lamp is often placed directly in front of the object is that of press photography. To be able to post himself easily at the best vantage point and to take his photographs quickly in any situation arising, a press photographer usually mounts his lamp (in this case it is always a flashbulb) on the camera itself. The same principle has lately gained a remarkable popularity also in amateur photography on account of the great advantage it offers in being relieved of all the paraphernalia of stands and electric leads (and thinking about the right positioning) for taking snaps in the home. In fact cameras are now on the market with a built-in reflector for a flashbulb; an example is the Philips flash camera [6]) illustrated in *fig. 5*.

Here it is to be noted that with the lamp fixed on the camera the requirement of distance derived above with respect to the width of the object is automatically satisfied. In fact the camera has only a limited field of vision. When the source of light is in the same position as the camera those parts of the object that are too far removed to the side to be properly illuminated do not appear in the photograph anyhow. Owing to lens aberrations, with a normal photographic objective only a part of the focal plane is used having a diameter

[4]) J. A. M. van Liempt, Kunstlicht in de fotografie (Artificial light in photography; in Dutch), Philips Techn. Library 1942, p. 113.

[5]) The results of these considerations have in part already been used in a publication in "Photographische Korrespondenz" 80, p. 101.

[6]) Here the lamp is mounted immediately above the lens. When working with only one lamp this position is relatively the best from the point of view of shadow formation; if the shadows cannot be softened then it is better not to have much shadow in the picture, and it is just with the lamp in this position that the eye of the camera sees almost nothing of the shadows cast by the lamp.

Fig. 5. Philips flash camera, type No. 5210. It contains a built-in reflector for a flashbulb. The camera is also fitted with a synchronizer, a pilot lamp for the lamp circuit, a plug socket for connecting additional flashbulb, and a compartment to take the battery for the ignition.

approximately equal to the focal length. This means, roughly, that $r = 2l$, thus $K = 1.25$ for non-planar objects [7]).

When large-angle objectives are used the foregoing no longer applies. These objectives have much larger fields of vision, corresponding, for instance, to an angle of 120°, thus $r/2l = 0.29$. If in this case the lamp were to be mounted on the camera then in the width of the object there would be an inequality factor $K = 4$ for non-planar and $K = 8$ for plane objects — apart from the effect of other causes of inequality, which will be mentioned below and which have to be taken all the more into account with such large angles.

Let us now consider the depth of the object. In this respect, too, the distance requirement derived above will automatically be satisfied, within certain limits, when the lamp is mounted on the camera, since the "field of vision" of the camera is limited also in depth, if we consider the area that the camera is able to focus. The deeper the object, the farther the camera has to be away from it in order to encompass the whole with its "focus depth". It is, however, to be noted that the ratio r/d required on that account is only greater than that according to (2) in the case of close-ups and/or when using a large relative aperture.

[7]) Here the "width" has been measured in the diagonal of the picture.

With some simplifying suppositions it can be deduced that the focus depth amounts to

$$d = \frac{2\,u\,m}{f^2}\,r^2\,,$$

where u represents the permissible blurring of the picture (diameter of the spot on the film in which a point of the object is pictured), f is the focal length of the objective and m = diaphragm stop (= f divided by the diameter of the objective). Taking $u \leqq f/1000$, as is mostly done for good objectives, we find:

$$\frac{r}{d} \geqq \sqrt{\frac{500f}{m\,d}} \quad \ldots\ldots\ldots\ldots \quad (5)$$

It is seen that m and d must not be too large, otherwise the minimum value of r/d required for focusing will not be large enough to ensure uniform lighting. Equation (5) is graphically represented in *fig. 6*, where a K-scale has also been drawn according to the relation given in fig. 3. To take a numerical example: if the lamp is mounted on a camera with f = 11 cm and a diaphragm stop $m = 7$ is used, then up to object depths of about 40 cm the requirement of sufficient focus depth automatically precludes any inequality in the illumination greater than $K = 1.5$. Or, to put it in other words, with these values of f and m sufficient uniformity in the illumination of the whole of the area covered by the focus depth is only ensured when the photograph is taken from a distance less than 1.80 m.

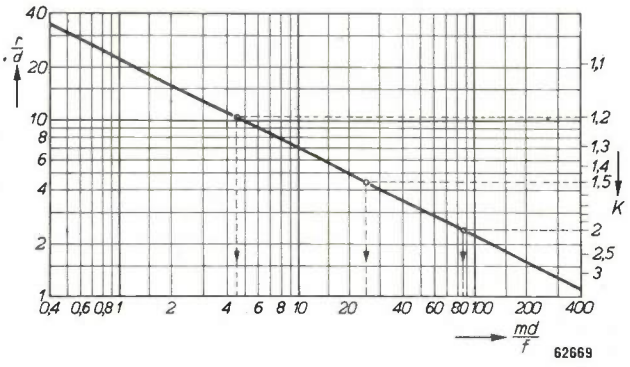


Fig. 6. To get the focus depth d the camera has to be placed at a distance r from the object. In accordance with the approximated equation (5) r/d is plotted as function of md/f (f = focal length, m = diaphragm stop). The scale on the right gives the values K appertaining to r/d according to fig. 3.

Arrangement of given objects

As a consequence of the conditions indicated in respect of uniformity, if the light from the lamp is to be used as economically as possible then a given number of objects or persons that can be placed side by side or one behind the other have to be arranged by the photographer in a certain manner. Let us assume, for the sake of simplicity, that when arranged in any order the objects cover the same area A; the considerations will be confined to the case where the objects are arranged in the form of a rectangle with width $2l$, depth d; $2\,ld = A$. If a very wide and shallow rectangle is chosen the lamp has to be placed far away on account of the condition (3); if the rectangle is made narrow and deep then the lamp has again to be at a great distance owing to the condition (2). The required distance r will have a minimum somewhere between the

two. With the aid of *fig. 7a* we find the condition for sufficient uniformity of illumination ($K = 1.5$):

$$l^2 + (d + r)^2 = 1.5\, r^2 .$$

Hence:

$$r = 2d + \sqrt{6d^2 + \frac{A^2}{2d^2}}, \quad \ldots\ldots \quad (6)$$

and taking the differential quotient of r and d as being 0 we get for the depth d of the rectangle where r is a minimum:

$$d = 0.425\,\sqrt{A} .$$

The most favourable rectangular arrangement therefore is that where

$$2l = A/d \approx 5.7\, d ,$$

while the lamp then has to be at a distance

$$r \approx 2.8\,\sqrt{A} .$$

The minimum of r is not strongly pronounced, as may be seen when $r/\sqrt{A}$ according to (6) is plotted as a function of $d/\sqrt{A}$; see fig. 7*b*. The ratio of the width to the depth of the rectangle, likewise indicated along the abscissa, may vary between 3 : 1 and 10 : 1 without the required distance having to be much greater than the minimum. If, therefore, a group of, say, 24 people is to be photographed (taking for each person a circle as base) then, as far as the illumination is concerned, it matters little whether they are arranged in two rows of 12 or in three rows of 8, but it would be less satisfactory to arrange them in one row of 24 or in four rows of 6.

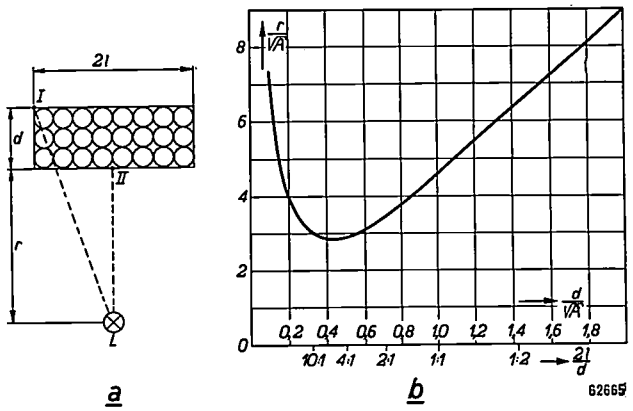


Fig. 7. *a*) A given number of objects are arranged in a rectangle. The point *I* receives the smallest and the point *II* the greatest illumination.
b) With a constant area A of the rectangle, for any shape (width $2l$, depth d) a certain minimum distance r from the source of light is required to get sufficient uniformity of the illumination. As function of the ratio $2l{:}d$ of the sides of the rectangle this distance shows a minimum at $2l \approx 5.7d$. Plotted in this diagram are the dimensionless quantities $r/\sqrt{A}$ and $d/\sqrt{A}$.

Other factors affecting the uniformity of the picture

Even though the illumination of the object in its width may be absolutely uniform, on the negative the density will still diminish from the centre towards the edges. This is due to similar causes as discussed in the foregoing in respect of the object: the distance from the "source of light" (in this case the objective) to the edges is greater than that to the centre of the photographic plate or film, and the parts near the edge stand at an angle to the rays of incident light; moreover the effective diameter of the objective is smaller for an oblique pencil of rays. These effects, taken together, cause the density to decrease as $\cos^4\alpha$ (α = angle between the ray of light and the normal of the plate). With normal objectives the luminous intensity at the edges of the plate is consequently 25% less, and in the case of large-angle objectives even as much as 75% less, than that in the middle of the plate. Here no account has yet been taken of the effect of possible vignetting (partial shading of the diaphragm aperture by the lens mount of parts of the objective closer to the plate), nor of the fact that the light falling on the plate is not fully used owing to part of it being reflected back from the surface of the plate — two effects that become more and more pronounced the larger the angle α.

Of course the decrease in density of the negative towards the edges due to all these effects mentioned occurs just as well in photographs taken by daylight as it does in those taken by artificial light. It can only be avoided by limiting the angle α (i.e. using cameras with a reasonably large focal length), as is strongly recommended for colour photography and is the general practice in reproduction work. The results of our considerations regarding the illumination in normal cases can only be affected by this phenomenon in so far that the permissible differences in illumination between the middle and the edge of the object are further decreased, so that even greater distances from the source of light are required.

So far it has been assumed that the luminous intensity of the lamp is the same in all directions. In practice, however, this is never the case, because reflectors are always used and these more or less concentrate the light. Usually the luminous intensity within the solid angle determined by the object is not entirely constant, but shows a slight decrease — once more — towards the edges of the object; this may at times have to be taken into account.

Reflectors may also be used which give a directional distribution of the light bearing just the opposite character, the luminous flux increasing to a certain extent from the axis towards the edge of the beam ("broad radiators", such as the "Comptalux" lamp). In that case the aforementioned reduction of the illumination towards the edge of the object, and the analogous decrease in intensity of exposure of the plate, is more or less compensated and the lamp can be placed much

closer to the object — closer than the camera! But then the strong reflexes and cast shadows appearing round the edges of the object may prove to be a drawback: owing to their being differently directed on opposite sides of the object these cast shadows (which are now not invisible, cf. footnote [6])) give the picture an unnatural impression.

It will presently be seen that the decrease in intensity referred to can also be compensated when using more than than one lamp as the main light.

Mention has already been made in the introduction of the equalizing effect of the diffuse supplementary light. The greater the contribution from this lighting, the more it is permissible to deviate from the conditions given for the positioning of the main light.

Finally, there is a very important factor that we have left out of consideration in all this. What counts ultimately is the difference in density in the positive, and this can be greatly influenced by the manner in which the negative is developed, or by the way in which prints are taken from it or enlargements made. When making an enlargement the photographer can print on the relatively too strongly exposed middle part by locally screening off the light. By this artifice he cannot, however, in general compensate inequalities of the illumination in the depth of the object, nor can he thereby save any parts of the negative that have been so under- or over-exposed, due to inequality of the lighting, that the details have been lost in the flat parts of the density curve of the negative.

A single lamp to the side facing the object

We shall now consider the case where one single lamp is set up not directly facing the object but to the side in front of it (some of the remarks in the previous section hold equally, of course, also for this set-up). Let the perpendicular distance to the object plane be r and the distance to the central plane perpendicular to the object s (*figs 8a* and *b*).

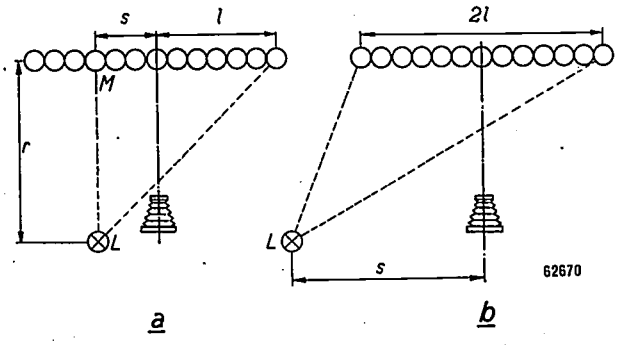


Fig. 8. A source of light L placed at a distance s from the central plane perpendicular to the object. *a*) $s < l$; *b*) $s > l$.

If $s < l$ then the part M immediately facing the lamp receives the greatest illumination, whilst the illumination is a minimum at the edge of the object at a distance $s + l$ from M. Thus the situation is analogous to that where a lamp is set up centrally in front of the object and we can again apply equations (3) and (4), only now instead of $2l$ we have to take $2(s + l)$ as the width of the object.

If the lamp is placed still farther to the side, so that $s > l$ (fig. 8*b*), then the best illuminated part lies at the nearer edge of the object and the least illuminated part at the farther edge. The ratio between the two degrees of illumination is:

$$K = \frac{r^2 + (s+l)^2}{r^2 + (s-l)^2} \quad \text{or} \quad K = \left[\frac{r^2 + (s+l)^2}{r^2 + (s-l)^2}\right]^{3/2}$$

respectively for a group and for a plane object. From this we derive the condition:

$$\frac{r}{l} \geqq \frac{1}{\sqrt{G-1}} \sqrt{\left(\frac{s}{l}+1\right)^2 - G\left(\frac{s}{l}-1\right)^2}, \quad (s > l) \quad (7)$$

where, for a group, G = the permissible maximum value K, and for a plane object $G = K^{2/3}$.

For any lateral distance s we can therefore calculate the minimum perpendicular distance r required. When we draw a ground plan of the space in which the object and the lamp are situated then all points yielding exactly the permissible value for K can be connected by a line, and this way we separate "prohibited" from "permitted" areas for the placing of the lamp. This has been constructed in *figs 9a* and *b* for the two cases of a group and a plane object. (The figure can be used for any width of object $2l$, since the conditions derived contain only the dimensionless quantities r/l and s/l.) The hatched part is free, the white prohibited. The dotted lines indicate the limits which would be obtained if the permissible difference in illumination were taken as a factor 2 or 1.2 respectively instead of 1.5.

What is most striking in these ground plans is the great economic advantage of placing the lamp centrally in front of the object ($s = 0$) as compared with a position to the side of it. When $s = 3l$, for a group, a luminous intensity about 12 times as large is needed to get the same average illumination of the object as when $s = 0$. It is true, however, that when $s = 3l$ only a much smaller angle of the light cone of the lamp is used, and we could therefore use a lamp fixture with stronger concentration of the beam (a reflector with a greater amplifying factor).

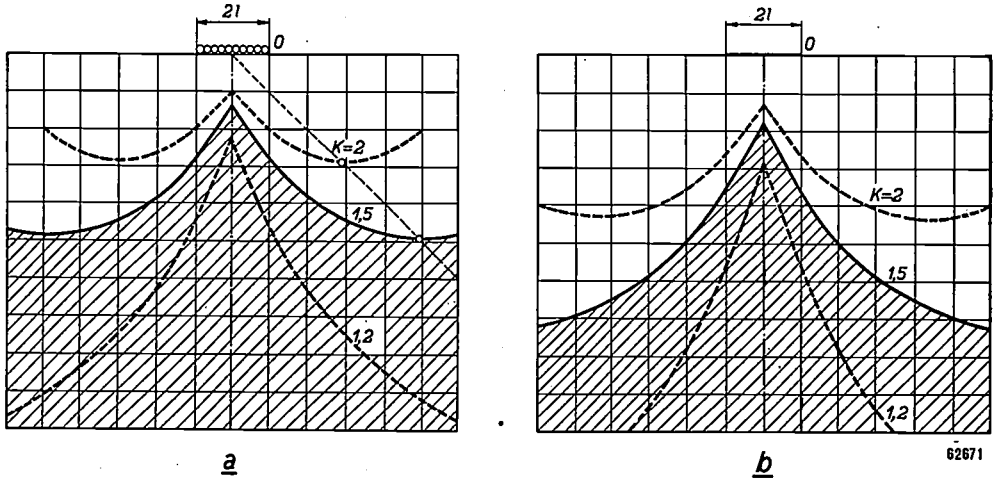


Fig. 9. Ground plans of probihited and permitted areas (the latter hatched) for placing a source of light with respect to a given object, drawn to scale, with a permissible inequality $K = 1.5$. *a*) The object consists of a group of people placed in a row. *b*) The object is plane (reproduction).
The dotted 45° line refers to the Mohler arrangement (see in the text below).

As a special case in this connection mention may be made of the frequently applied Mohler method of illumination sketched in *fig. 10*. A powerful lamp (main light) is set up at the point $r = {}^1/_2 R$, $s = {}^1/_2 R$, where R is the distance from camera to object, and a less powerful lamp (supplementary light) is set up close to the camera. A rough calculation shows that when the latter lamp has, say, half the strength of the primary one this supplementary light does not appreciably improve the uniformity of the illumination (of course it gives the desired softening of the shadows). With the aid of the ground plan in fig. 9*a* and the dotted 45° line, and given the distance R from camera to object, it is possible to deduce directly the permissible width of the object for this method of illumination. From the same figure (imagining the camera to be set up in the produced object line) we can also find the permissible depth, which proves to be about equal to the permissible width. We find:

$K =$	1.2	1.5	2
$\frac{d}{R} \approx \frac{2l}{R} \leqq$	0.09	0.21	0.37

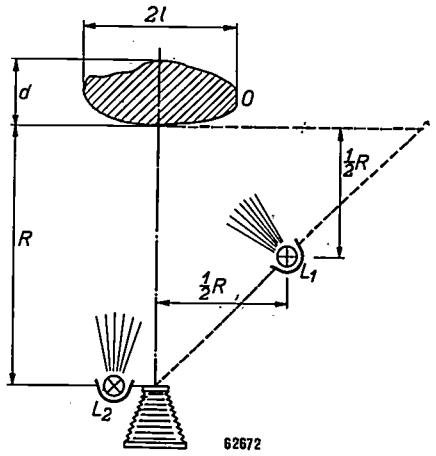


Fig. 10. The arrangement according to the Mohler method of illumination. L_1 main light, L_2 supplementary light, O object.

Two lamps set up symmetrically

The arrangement of two lamps symmetrically with respect to the central plane perpendicular to the object (*fig. 11*) is typical for reproduction work. For three-dimensional objects, with equal

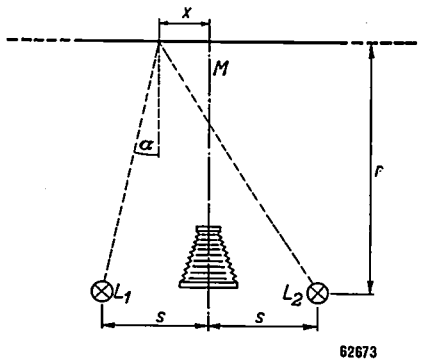


Fig. 11. Symmetrical set-up of two lamps L_1 and L_2 in front of a plane object. M is the plane perpendicular to the centre of the object.

contributions of light from two lamps such an illumination, (not standing close together), has several drawbacks, one of which is that it produces double shadows. For plane objects, on the other hand, there are no such objections and the arrangement even has distinct advantages: the inequality in the lighting can be minimized more economically than is possible with only one lamp.

The improvement that can be reached may be illustrated by constructing a ground plan similar to that in fig. 9*b*, but before doing so we wish to point out the advisability of including now in our considerations also linear sources of light. In normal photography linear sources of light are

used only for diffuse supplementary lighting and never for the main lighting, owing to the peculiar shadow formation [8]). With the plane objects that are photographed in reproduction work however, where this objection does not hold, use is gladly made of the advantages offered by the linear source of light; in the direction parallel to the source the illumination of the object is absolutely uniform (provided the length of the source of light is greater than the distance and the dimensions of the object), and in the direction at right angles to the source the decrease in the illumination with increasing distance R from the source of light follows only the relation $1/R$ and not $1/R^2$ as in the case of a point source of light (see the article quoted in footnote [8])).

This is in fact the principle mostly applied in the apparatus employed for reproduction work, using either linear sources of light direct (*fig. 12*) or a row of point sources of light forming together approximately a continuous line (*fig. 13*). The set-up sketched in fig. 11 will therefore be considered not only with two identical point sources of light but also with two identical linear sources of light (at right angles to the plane of the drawing). For the illumination produced by a linear source of light we have to use, instead of (1a), the formula:

$$E = \frac{I_l}{R} \cos \alpha, \quad \text{(1b)}$$

where I_l is defined as the luminous intensity per unit of length of the lamp.

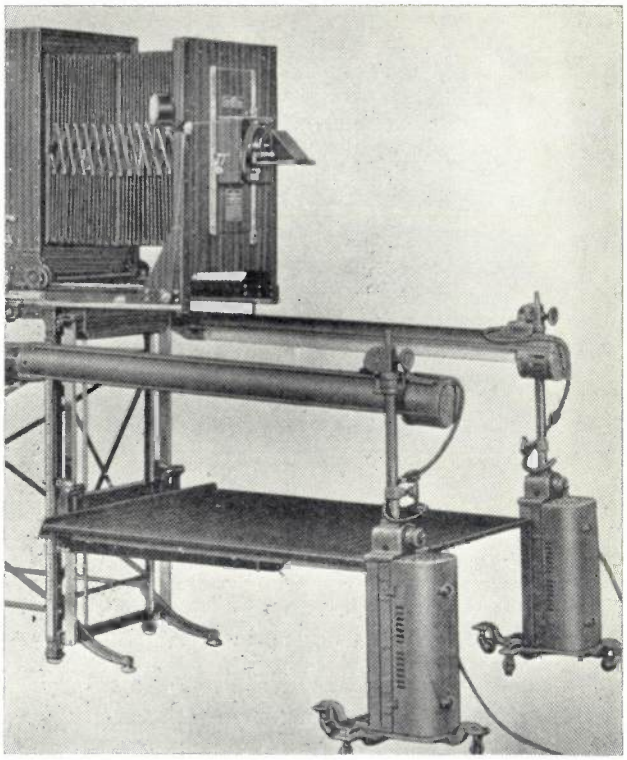
Contrary to the cases previously dealt with, with the two-lamp arrangement it cannot be seen at once in what points of the object plane the smallest and the greatest illumination occurs. To determine this, graphical means are employed. For the illumination in a point of the object at a distance x from the central perpendicular plane of the whole set-up (fig. 11) we have for point sources of light:

$$E = \frac{I}{r^2\left[1+\left(\frac{s}{r}-\frac{x}{r}\right)^2\right]^{3/2}} + \frac{I}{r^2\left[1+\left(\frac{s}{r}+\frac{x}{r}\right)^2\right]^{3/2}} \quad \text{(8a)}$$

and for linear sources:

$$E = \frac{I_l}{r\left[1+\left(\frac{s}{r}-\frac{x}{r}\right)^2\right]} + \frac{I_l}{r\left[1+\left(\frac{s}{r}+\frac{x}{r}\right)^2\right]} \quad \text{(8b)}$$

We now plot $E \cdot r^2/I$ and $E \cdot r/I_l$ (the factors r^2/I and r/I_l affect only the level of the illumination and not its distribution) as functions of the place coordinate x/r in the object plane, with s/r as parameter describing the set-up of the lamps. In this way we get for different set-ups the distribution curves produced in *fig. 14*.



62735

Fig. 12. The "Photostat" reproduction apparatus (manufactured by Photostat Corporation, Inc., Rochester, U.S.A.). The plate or page of a book to be reproduced lies horizontal and is illuminated with two linear lamps (mercury lamps 1.20 m long in enamelled reflectors). In a simpler design, the "Photostat Junior", two incandescent lamps are used, thus more or less point sources of light, placed in the same way. — In a reproduction apparatus the object is photographed via a prism placed in front of the objective, thereby inverting the image. This is necessary to allow reproductions to be made direct on "positive" paper.

For each curve we ascertain up to what coordinate x/r the differences in illumination do not exceed a factor $K = 2$, or 1.5, or 1.2 (the value $K = 2$ is far too large for reproduction work but it has also been included again for the sake of completeness). In *fig. 15* the values found, x_{max}/r, have been plotted as a function of s/r. From these curves we derive directly the permissible width of the object (x_{max}) with a given set-up of the lamps (s and r), and inversely a relation between r and s can be derived when the object width $2l$ ($=2x_{max}$) is given. This relation, which is the ground plan of the prohibited and the permitted areas for the symmetrically arranged point sources of light or linear lamps, is shown in *figs 16a* and *b* respectively.

It is seen that the boundary curves in the ground plan show a discontinuity on each side. This

[8]) Cf. N. A. Halbertsma and G. P. Ittmann, Illumination by means of linear sources of light, Philips Techn. Review 4, 181-188, 1939.

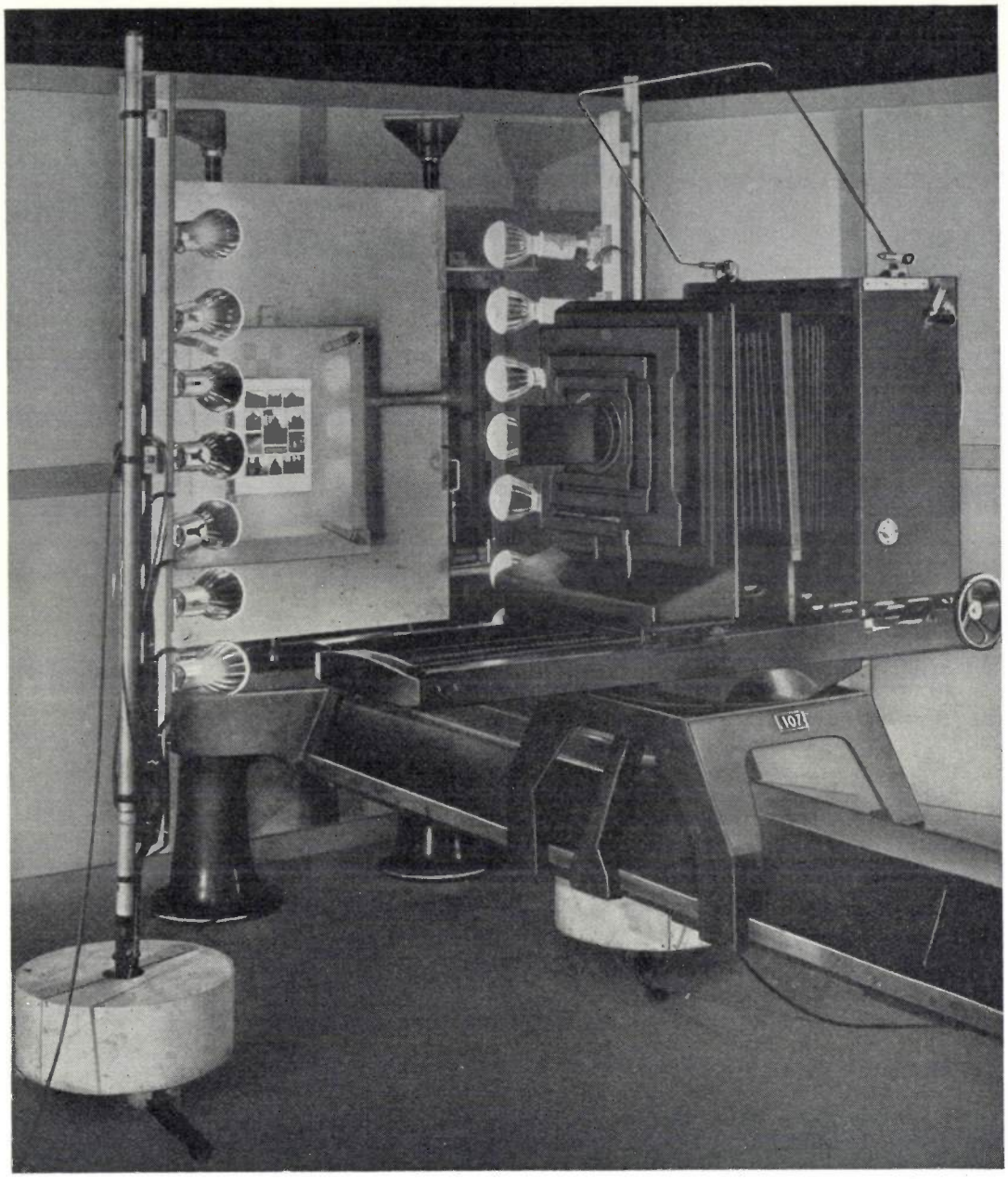


Fig. 13. Reproduction apparatus with the object illuminated by means of two vertical rows of "point" lamps (high-pressure mercury lamps in special reflectors, type HPR). (This photograph has been taken from the article by Guinau quoted in footnote [9].)

corresponds to the jump in the curves of fig. 15 for a certain value of s/r. The significance of this becomes clear when we look closer at fig. 14. The dip in the middle of the curves is seen to become deeper the larger the value of s/r, and when a certain s/r is reached the illumination in the middle becomes exactly $1/K$ times the illumination in the two maxima. At this value of s/r the object width may extend a certain distance beyond the two maxima, but with the slightest further increase of s/r the width suddenly becomes limited to between the two maxima (and as s/r still further increases the width is confined to a decreasing part of this central area). Whereas, therefore, with a small s/r we work with the illumination distributed over the object as sketched in *fig. 17a*, with a large s/r we have to work with a distribution according to fig. 17*b*. It is seen that here we have the possibility of compensating more or less the decrease of illumination towards the edges of the negative that is inherent in the photographic camera.

The case $s = 0$ is of course identical to that of a single lamp placed centrally in front of the object. Point A in fig. 16*a* therefore answers the condition

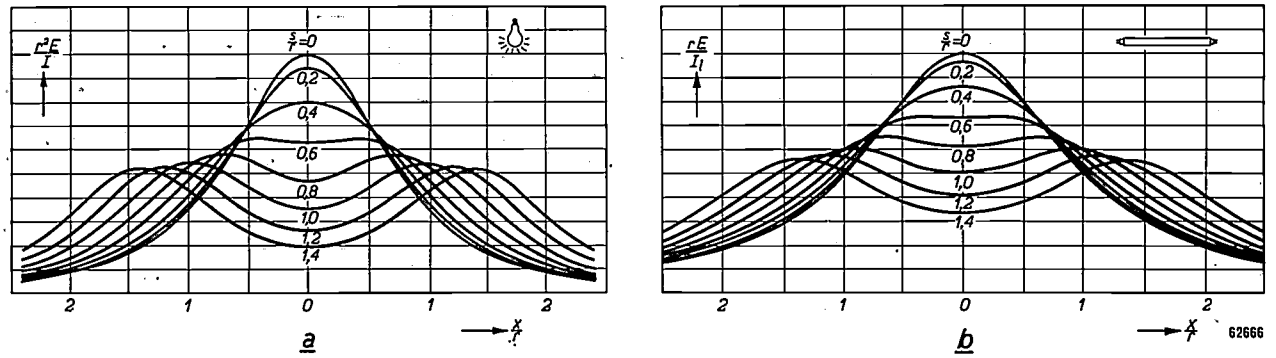


Fig. 14. Illumination (on a relative scale) as a function of the point x in the object plane with two sources of light arranged symmetrically according to fig. 11 at a mutual distance $2s$,
a) for two point sources of light; *b*) for two linear sources of light.
The distribution curve for $s/r = 0$ is the same as with one lamp placed centrally in front of the object. With increasing s/r the maximum becomes flatter, for $s/r > 0.6$ it splits up into two maxima lying at $x \approx s$, thus approximately facing each lamp.

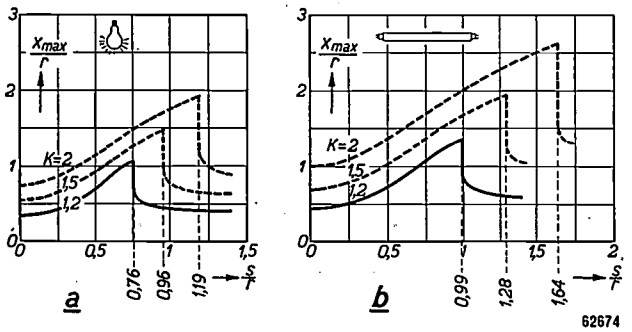


Fig. 15. For each curve (s/r) of fig. 14 the abscissa (x/r) is sought for which the illumination is a factor $K = 2$, 1.5 or 1.2 less than in the maximum (maxima) or greater than in the minimum. The values found, x_{max}/r, are plotted here as functions of s/r:
a) for two point sources of light,
b) for two linear sources of light.

of equation (4). The discontinuity of the boundary curve for the same K in fig. 9*b* (boundary case of one lamp to the side facing the object) likewise falls in this point A; this curve is indicated in fig. 16*a* by a thin broken line. The advantage of using two lamps is obvious. A comparison of the ground plan in fig. 16*b* with that in fig. 16*a* demonstrates the advantage of the linear sources of light.

As regards the further choice of position of the lamps in figs 16*a* and *b*, for the most economical use to be made of the lamps — in reproduction work this is a particularly important aspect — the best would be round about the discontinuities the curves, but then the lamps would be in the field of vision of the camera [9]). Moreover, account has also to be taken of the disturbing

[9]) For making "contact prints", such as are used in the production of autotype prints, where the negative or diapositive is provided with a raster (blocks, offset plates), no camera is used, so that two lamps can then quite well be placed at the discontinuitics. However, for this process the system of one point source of light (a carbon arc lamp) placed centrally in front of the object is still commonly used, apparently because of difficulties in rendering the raster points owing to the Callier effect. It goes without saying that this arrangement, in the present case, where even an inequality factor of $K = 1.2$ is still too large, is very inefficient. O. A. Guinau has recommended, for this case, a series of lamps filling a square plane parallel to the plate (International Lighting Review 1949/50, No. 1, p. 26).

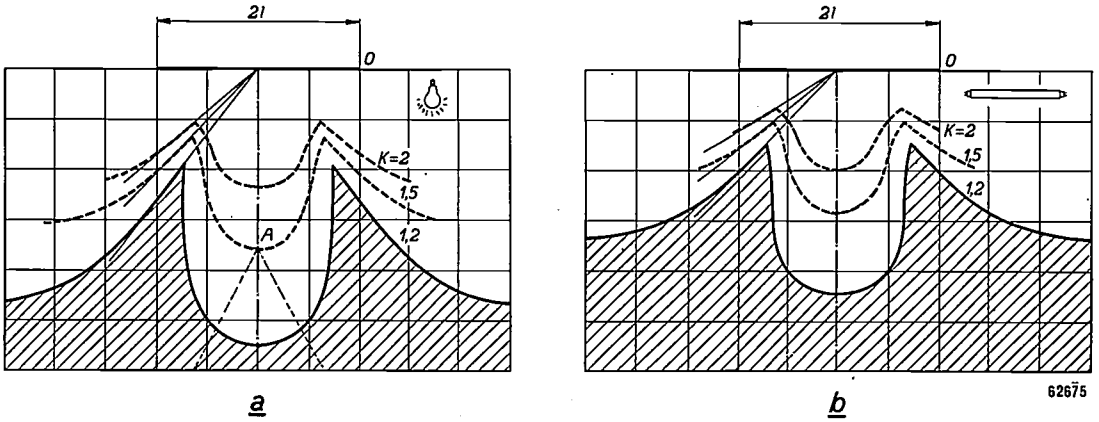


Fig. 16. Ground plan of prohibited and permitted areas (the latter hatched) for the symmetrical placing of two lamps, with a permissible inequality factor $K = 1.2$ in the illumination. The boundary lines for $K = 2$ and $K = 1.5$ are dotted.
a) For two point sources of light, *b*) for two linear lamps.

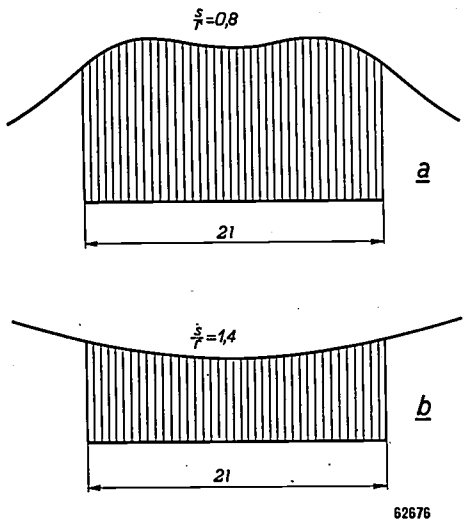


Fig. 17. Variation of the illumination over the width of the object.
a) with two lamps set up in the central area of the boundary curves in fig. 16; e.g. at $s/r = 0.8$ on the curve $K = 1.2$ in fig. 16*b*;
b) with two lamps set up on the "edge" parts of the curves (beyond the discontinuities); e.g. at $s/r = 1.4$ on the same curve as for *a*.

reflexions apt to arise when reproducing objects tending to shine. To avoid these difficulties the lamps have to set be up outside the border lines g drawn in *fig. 18*. By a combination with fig. 16 a permitted area is left for the position of the lamps as indicated by the hatched parts in fig. 18. The ideal position is then in the points P.

When two linear sources of light are used, arranged on either side of the object, the uniformity of the illumination can obviously be still further improved by adding two identical horizontal linear lamps in the places analogous to P underneath and over the object. The inequality factor is then reduced from $K = 1.2$ to $K = 1.1$. At the same time, it is true, an inequality is introduced in the vertical direction, which — if the linear lamps are of sufficient length — does not otherwise occur, but this inequality too is only a factor 1.1. In this way it is therefore possible to meet, with relatively simple means, very high requirements with regard to the uniformity of the illumination.

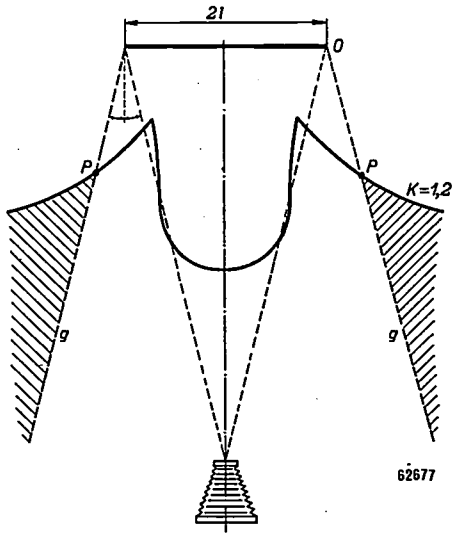


Fig. 18. In order to avoid disturbing reflexions the lamps have to be placed outside the boundary lines g, g (which depend upon the distance from the camera). When this condition is combined with the uniformity requirement $K = 1.2$ the hatched part remains as the permitted area for the lamps.

Summary. When photographing with artificial light the illumination of the object shows a certain inequality, owing to the relatively short distances between the object and the sources of light. When taking a certain permissible value, say 1.5, for the ratio of the greatest to the least illumination a number of simple rules can be formulated for the positioning of the lamps (of course this does not yet take into consideration any artistic requirements). This has been done and in this article it is explained how these rules are applied for the case where only one lamp is used, whether or not in the central plane perpendicular to the object, and for the case where two point sources of light or linear lamps are arranged symmetrically. The results are given partly in the form of ground plans on which "prohibited" and "permitted" areas for the lamps are marked off.

VOL. 12 No. 6, pp. 157-184 DECEMBER 1950

Philips Technical Review

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

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

TWO TRANSMITTING VALVES FOR USE IN MOBILE INSTALLATIONS

by E. G. DORGELO and P. ZIJLSTRA. 621.385.4:621.396.61:621.396.72

When in 1896 Marconi succeeded in developing wireless telegraphy into a technically useful means of communication the main application envisaged was for establishing communication between ships at sea and the shore. In those days there was little difference between a fixed radio station and a mobile one, but this changed when, some time later, wireless came to be applied also to aircraft and equipment began to appear in special forms most suited for their purpose. It was about that time, too, that telegraphy was largely ousted by telephony. It is only in recent years that radiotelephony has come to be used on a large scale in automobiles and trains, for which purpose special frequency bands were allotted by the last radio conference at Atlantic City.

As regards the transmitting valves developed for this new form of mobile installations, they are required to possess great mechanical strength combined with high efficiency, the latter property being demanded on account of the fact that the most efficient use possible has to be made of the power available, since this is supplied by a source of only limited capacity, in a car, for instance, by the battery.

In recent years small, mobile, radio-telephonic transmitting and receiving installations have become very popular as a means of meeting a need felt in many directions. When automobiles, trains or small vessels are equipped with these sets communication can be maintained between them and with one or more fixed stations. As examples may be mentioned: taxi companies, fire brigades, police forces, military columns, harbour works, railways (both for the convenience of passengers and for use in marshalling yards), outposts or remote plantations in inhospitable regions, doctors and business people desiring to keep in touch with their homes, offices or works, etc., etc.

For such purposes as these the transmitter and the receiver are usually built together, sometimes making such a compact whole that it can be carried on the back or even in the hand (we have in mind the "walkie-talkies" and the "handie-talkies" which rendered such good services in the allied armies during the war).

This article deals with two transmitting valves that have been developed specially for such mobile installations. The need for new valves arises mainly from the fact that the supply source — the accumulator battery of a car or the dry batteries of the smaller, portable sets — is of a very much more limited capacity than in the larger mobile installations. It is therefore necessary that the transmitting valves should work with a high efficiency and possess such properties that a small number of stages suffices (for amplification and frequency multiplication). Furthermore, valves destined for use in trains or automobiles have to withstand greater mechanical shocks than occur on board ships.

Before proceeding to describe these new valves it is deemed necessary to consider briefly some points that are of importance for mobile transmitters.

Mobile transmitters

Wavelengths

Mobile radio stations were officially recognised at the international telecommunication conference held at Atlantic City in 1947, when a number of frequency bands were allotted to these stations.

For such of these bands as lie between 54 and 420 Mc/s (5.55 to 0.714 m waves) their distribution is indicated in *fig. 1*.

The other bands outside these limits have not been included in this diagram for the following reasons. At frequencies below a certain limit there is a risk of the waves being reflected by the ionosphere and reaching the earth again in parts far removed from the transmitter [1]); welcome use is made of this for long-distance radio communication, but for transmitters intended only for short-range work, like mobile transmitters, this reflection by the ionosphere is undesirable because it is apt to lead to interference over wide ranges where the same wavelengths are being used. At frequencies of 54 Mc/s and higher there is no need to take account of this effect, although at the highest allotted frequencies under certain atmospheric conditions an inversion of the temperature gradient occurs and the signals may cover a range of some hundreds of kilometres, but then the ionosphere has nothing to do with this; such conditions, however, very seldom occur.

The other limit, 420 Mc/s, is about the highest frequency at which the transmitting valves to be described here still work with a satisfactory efficiency; for still higher frequencies entirely different types of valves would be required.

For the allocation of frequencies for working mobile transmitters the world has been divided into three zones, where the following arrangements apply for the frequencies between 54 and 420 Mc/s.

In *zone I*, covering Europe, Africa and some parts of Asia, the frequency bands that may be used are: 70-72.8, 75.2-78, 80-83, 85-87.5, 100-108, 156-174, 235-328.6 and 335.4-420 Mc/s.

There are some exceptions however: the 85-87.5 Mc/s band does not apply for Great Britain, which has instead the disposal of the 66.5-68 Mc/s band; South Africa and its mandatory territories have been allotted, instead of the 100-108 Mc/s band, the bands 133-144 and 146-174 Mc/s and also, but only for mobile stations used for broadcasting, the 54-68 Mc/s band.

Zone II, comprising mainly North and South America, has been allotted the 54-88, 132-144, 148-220, 225-328.6 and 335.4-420 Mc/s bands.

Zone III, consisting mainly of that part of Asia not included in zone I and of Australia and New Zealand, has the disposal of the bands 54-68, 70-78, 80-87 (in Australia and New Zealand 80-85), 132-144 (not in Australia and New Zealand), 148-200 (in Australia limited to 156-170 and 178-200, in New Zealand limited to 156-200), 235-328.6 and 335.4-420 Mc/s.

In order to avoid interference with or from other radio communications, mobile transmitters are not allowed to work outside the fixed frequency bands, and in order to avoid any mutual interference their frequency has to be highly constant. For this reason crystal control has to be applied.

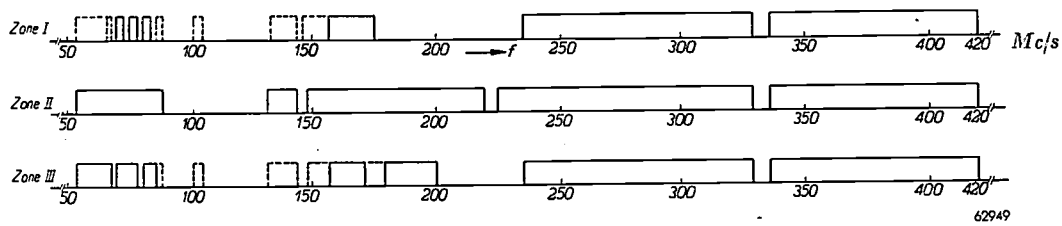


Fig. 1. The frequency bands allotted to mobile transmitters by the International Telecommunications Conference held at Atlantic City in 1947, for so far as these bands lie between 54 and 420 Mc/s. Roughly speaking zone I comprises Europe, Africa and part of Asia, zone II North and South America, and zone III the part of Asia not belonging to zone I, and Australia and New Zealand. The bands denoted by dotted lines can only be used in certain parts of the respective zone. For further details see the text in small type.

Power and range

The power of mobile transmitters is closely related on the one hand to the minimum range required and on the other hand to the permissible weight and volume of the apparatus, including the power supply unit.

The range depends not only upon the power radiated by the aerial but also to a large extent upon the construction and height of the aerial, whilst also the surroundings are of great influence. The aerial constructions most favourable for radiation are generally too cumbersome to be of use on an automobile or train. The higher the aerial the greater is the range, but then of course the height is limited for travelling vehicles; on an automobile, for instance, as a rule a simple vertical rod on the roof is used. In open spaces the range is much

[1]) See, e.g.: C. J. Bakker, Radio investigation of the ionosphere, Philips Techn. Rev. 8, 111-120, 1946.

greater than in built-up areas; the adverse influence of buildings is, roughly speaking, felt most at the highest frequencies, though there are exceptions to this rule. In New York, for instance, it has been found that in narrow streets shorter waves are more readily reflected downward by buildings than are longer ones. A second reason for shorter waves being sometimes more satisfactory in a town is that in a street standing waves are apt to be set up; when driving through such a street one therefore passes through maxima and minima, and the quicker these follow each other — the smaller the wavelength — the less does intelligibility suffer.

Intermittent working

If a call is to be heard at once, the receiver in a mobile radio station must be continuously in the stand-by position. After a call has been received, or when another station has to be called up, the installation is changed over to transmission, mostly by means of a switch built into the microphone handle. This changing over is then repeated as the conversation proceeds to and fro. Since as a rule the intervals of rest between talks are much longer than the talks themselves, in practice the transmitter is working only a fraction of the time, while the receiver is working almost continuously. Thanks to these long intervals, for some transmitting valves operating conditions are allowed which permit of a somewhat larger output than would be permissible for continuous working.

Number of stages of the transmitter

Notwithstanding the relatively short working time of the transmitter, it is necessary to aim at the least possible power consumption, i.e. at a high total efficiency of the transmitter. We shall presently deal further with the steps that have been taken in the new transmitting valves in order to limit the losses as far as possible, but another equally important factor upon which the total efficiency depends is the number of stages required between the aerial and the quartz crystal determining the carrier frequency.

This number of stages is closely related to the system of modulation. Two systems, amplitude modulation and frequency modulation, are to be considered, both of which are applied in mobile transmitters; it is not the place here to explain why in one case amplitude modulation is employed and frequency modulation in another.

Where amplitude modulation is employed one has a constant carrier frequency, which of course has to lie in one of the frequency bands allotted. Now the frequencies of these bands are so high that they cannot be generated directly by means of a crystal. Owing to the fragility of the crystal, especially when it is exposed to the shocks occurring in a mobile transmitter, about 0.15 mm is the least thickness it may have, and this corresponds to a natural frequency of about 20 Mc/s [2]). For a carrier frequency of say 320 Mc/s it is therefore necessary to apply at least a 16-fold frequency multiplication. In such cases it is advantageous to use double valves with the two electrode systems, mounted in one envelope, connected in cascade. By frequency doubling in each system no more than two of these double valves are then required to multiply the input (crystal) frequency 16 times.

In addition an output valve and a modulator valve are needed and one or two stages of A.F. amplification between the microphone and the modulator.

With frequency modulation a system can be followed, for instance, as described by Braak [3]), where the microphone voltage brings about a phase shift between two currents having the frequency of the quartz crystal, in such a way that the phase φ of the sum of these currents varies according to the equation

$$\varphi = \omega_0 t + \alpha \sin 2\pi\nu t$$

where ω_0 = angular frequency of the crystal, t = time, α = amplitude of the phase shift (phase sweep), ν = frequency of the incident sound at the microphone. Corresponding to the phase modulation of the total current is a frequency modulation, the instantaneous value ω of the varying angular frequency being defined by

$$\omega = \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \omega_0 + \alpha \cdot 2\pi\nu \cos 2\pi\nu t,$$

so that the instantaneous value $f = \omega/2\pi$ of the modulated frequency is:

$$f = f_0 + \alpha\, \nu \cos 2\pi\nu t$$

($f_0 = \omega_0/2\pi$). The sweep of this frequency is $\alpha\nu$. For undistorted reproduction the frequency sweep

---

[2]) Crystals are also being used, especially in the U.S.A. and in Great Britain, which oscillate with (practically) a multiple of the natural frequency (see, e.g., the article by W. Parrish in this number, p. 166), so that frequencies higher than 20 Mc/s can be generated with still crystals.

[3]) D. J. Braak, Mobile radio equipment, type SRR 192, Communication News **10**, 120-125, 1949 (No. 4).

has to be proportional to the intensity of the sound and independent of the audio frequency ν, which means to say that the phase sweep α has to be proportional to the sound intensity and inversely proportional to ν [4]).

If an n-fold frequency multiplication is applied between the modulating stage and the output stage the sweep of the aerial frequency will therefore amount to $n\alpha\nu$. For transmitters which are modulated only with speech — as is the case with mobile transmitters — this sweep has been limited by international agreement to 15,000 c/s, and in order to derive the utmost benefit from frequency modulation this sweep has to be used to the full. To prevent prohibitive non-linear distortion the phase sweep α has to be limited to an angle of about 0.5 radian. Hence the frequency multiplication required to get an aerial frequency sweep of 15,000 c/s with the maximum phase sweep, i.e. at the lowest audio frequency ν occurring in telephony (300 c/s), and with full phase modulation ($\alpha \approx 0.5$ radian), is

$$n \approx \frac{15{,}000}{0.5 \times 300} = 100 .$$

(Since this value of n is greater than that found in the case of amplitude modulation, with frequency modulation the mechanical strength of the crystal is not a determining factor.)

A frequency multiplication of about 100 can be obtained with three double valves, for instance as follows: $(3 \times 2) \times (2 \times 2) \times (2 \times 2) = 96$. For the first of these valves a low-power type suffices, which need not be made for particularly high frequencies, e.g. the double triode ECC 40, but for the third valve (and possibly also for the second one), in which the frequency is much higher than the crystal frequency, it is preferable to use one of the double tetrodes of the type QQC 04/15 or QQE 06/40 [5]) to be described in this article. The same types of valves can also be used as output valves (with either frequency or amplitude modulation), though then the two systems have to be connected in push-pull instead of in cascade; we shall revert to this later.

[4]) The latter can be reached to a sufficient approximation by means of a correcting network between the microphone and the phase modulator.

[5]) Here the letter Q means tetrode, QQ double tetrode, C directly-heated oxide-coated cathode, E indirectly-heated oxide-coated cathode. The figures 04 and 06 signify that the valve is intended for a supply voltage of 0.4 or 0.6 kV respectively, while the numbers 15 and 40 denote the order of the output in watts (as will presently be seen, in suitably chosen circuits these valves can yield much higher outputs than 15 or 40 W).

Choice of the type of valve

A tetrode, rather than a triode or a pentode, has been chosen because of a number of considerations, most of which apply in general to all transmitting valves for the frequency range in question, without being limited to valves for mobile transmitters. These considerations are the following.

At high frequencies a triode is in two respects at a disadvantage compared with a screen-grid valve (tetrode or pentode); in the first place a triode requires a greater driving power [6]) and thus also necessitates a larger number of amplifying stages, whilst in the second place with screen-grid valves, compared with triodes, it is possible to work with fairly high frequencies without (external) neutralization, i.e. compensation of undesired feedback such as arises, for instance, owing to the capacitance between anode and control grid.

That is why a screen-grid valve has been chosen. Of the two kinds to be considered, the tetrode and the pentode, the former is to be preferred at very high frequencies, since the absence of the third grid permits of a smaller anode capacitance, i.e. the capacitance between the anode and the other electrodes together. The absence of the third grid, however, makes it necessary to take certain steps for preventing any secondary electrons emitted by the anode reaching the screen grid; the purpose of the third grid (suppressor grid) in a pentode is to bring about between the anode and the screen grid a potential minimum suppressing the undesired secondary emission. Such a potential minimum, however, can be obtained by other means too, by giving a tetrode system such dimensions that, with the normal working currents and voltages, between the anode and the screen grid a concentration of space charge is brought about which provides the desired potential minimum [7]).

Furthermore, the secondary emission itself can be counteracted by coating the anode with a layer of a material from which the electrons do not easily emerge (see the article quoted in footnote [7])). Both these measures have been applied in the new types of transmitting valves.

The anode capacitance, already small owing to the absence of a third grid, can be still further reduced by dividing the electrode system into two

[6]) See J. P. Heyboer, Five-electrode transmitting valves (pentodes), Philips Techn. Rev. **2**, 257-265, 1937, in particular pp 260 and 261.

[7]) See, e.g.: J. L. H. Jonker, Secondary emission in output valves, Philips Techn. Rev. **10**, 346-351, 1948, fig. 3, curve *1*. The difficulty therein mentioned, which makes this method unsuitable for output valves of low-frequency amplifiers, applies in a much less degree for transmitting valves.

parts and connecting the two halves to the external circuits in push-pull. With given total dimensions of the electrodes (thus with a certain permissible heat dissipation in the valve) and certain distances between the electrodes, the input and output capacitances are then four times smaller, since the partial capacitances are in series instead of in parallel.

When the two electrode systems of the double valve obtained by this division are connected in cascade instead of in push-pull the same type of valve can advantageously be used also for frequency multiplying, as we have seen above.

Transmitting valves with two electrode systems in one envelope have in fact been known some fifteen years already. In the old designs the electrodes not carrying any high-frequency voltage (the cathodes and the screen grids) were connected in pairs by short wires or strips, and the centres (neutral points) of the interconnections were led out through the envelope, as were each of the two control grids and the two anodes. This is illustrated in *fig. 2a*. A difficulty arising with these valves was the self-inductance of the interconnection of the cathodes and of the screen grids (fig. 2*b*). At very high frequencies the influence of these self-inductances is not to be ignored. The self-inductance between the cathodes causes an undesired inverse feedback and constitutes a positive contribution towards the input damping [8]), so that in order to yield a certain output the valve needs a larger driving power. The influence of the self-inductance between the screen grids is manifested as a negative damping [9]), which is zero only at one certain frequency and at other frequencies may assume such a value that in order to avoid self-oscillation some form of neutralization or other has to be applied, especially for valves with a high mutual conductance. Below the frequency just referred to this neutralization can be brought about by introducing a capacitor of a certain value between each anode and the control grid belonging to the other anode. Above that frequency these capacitors have to be connected between each anode and its corresponding control grid.

How these complications have been avoided in the new designs of double tetrodes will be shown in the next section.

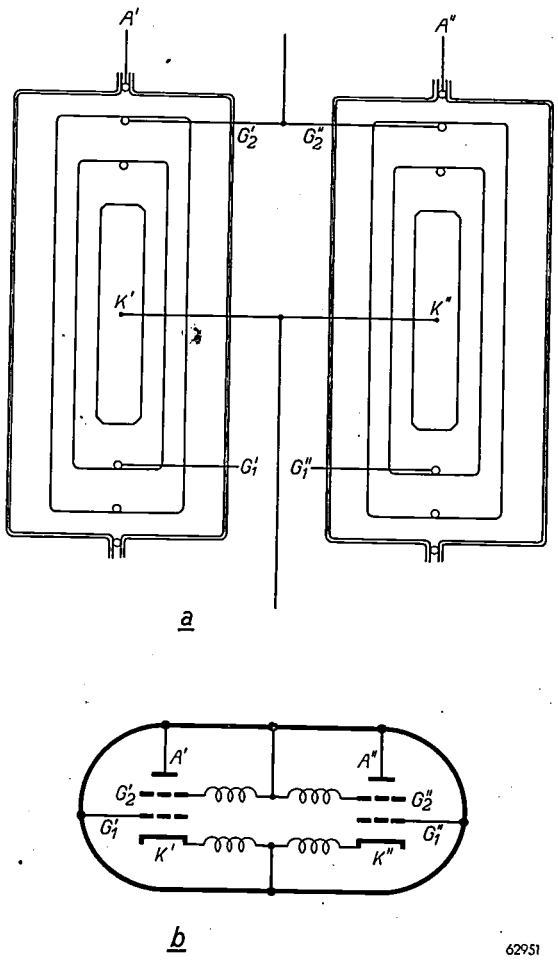


Fig. 2. (*a*) Cross section of a double tetrode of the old design. K', K'' cathodes; G_1', G_1'' control grids; G_2', G_2'' screen grids; A', A'' anodes. In the equivalent diagram (*b*) the stray self-inductances in the leads of the cathodes and screen grids are indicated. (The stray capacitances that are also present are not indicated.)

The double tetrode QQE 06/40

Construction

A double tetrode of the type QQE 06/40, illustrated in *fig. 3*, contains one indirectly heated, nickel cathode in the form of a roughly rectangular tube (*fig. 4*). Only the long, slightly convex sides of this tube are coated with an emitting material, so that really the tube has two cathodes interconnected by the shorter sides of the rectangular body. The self-inductance of these short and wide "connecting strips" connected in parallel is so small that even at frequencies of 400 Mc/s the aforementioned effect of self-inductance in the cathode interconnections is quite negligible. The resistance of this connection is likewise very small, even at high frequencies, due partly to the fact that the working temperature of the cathode lies above the Curie point of nickel, so that permeability is 1 and consequently there is but little skin effect.

[8]) M. J. O. Strutt and A. van der Ziel, A variable amplifier valve with double cathode connection suitable for metre waves, Philips Techn. Rev. **5**, 357-362, 1940.

[9]) W. G. Wagener, 500-Mc transmitting tetrode design considerations, Proc. Inst. Rad. Engrs. **38**, 611-619, 1948.

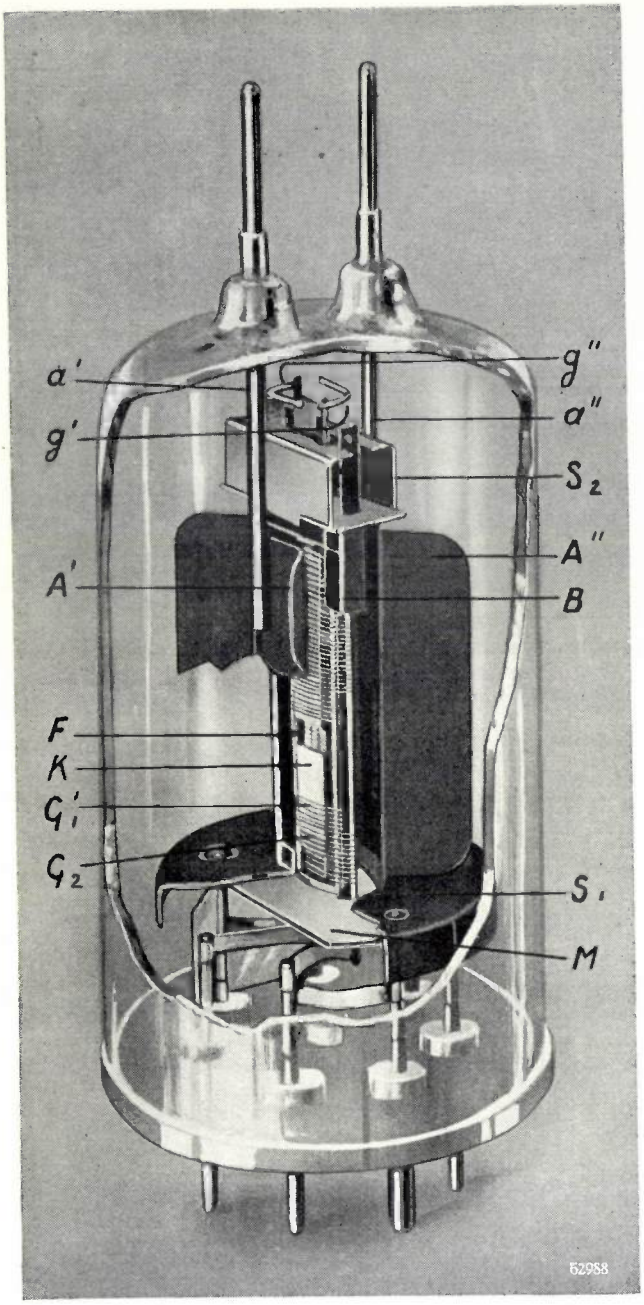


Fig. 3. Photograph of the double tetrode QQE 06/40 cut open to show the inside. K is one of the emitting cathode surfaces, F one of the filaments, G_1' one of the control grids, G_2 the screen grid, A', A'' the anodes, B one of the beam-plates, S_1 the lower screen screening the mica plate M in which the electrodes are fixed. The rods g', g'' are connected to the grids G_1' and G_1'' respectively and together with the anode poles a'' and a' form neutralizing capacitors. The box S_2 connected to the cathodes and the plates B screens the neutralizing capacitors from the electrode systems.

The cathode surface is heated by two filaments inside the cathode body.

A short distance away from and facing each of the emitting surfaces are the two control grids made in the form of a ladder. The extremely thin horizontal grid wires are curved so that when they expand the distance between the grid and the cathode is not reduced and thus there is no risk of short-circuiting.

The control grids are made of molybdenum wire plated with a thin layer of gold. This plating reduces the resistance at high frequencies and minimizes the risk of thermionic emission from the grid.

One single screen grid is placed around the system comprising the cathode and the two control grids. This screen grid is made of windings fixed to two supporting rods. This construction avoids the necessity of separate leads for the two halves of the screen grid and thus also completely eliminates the self-inductance of those leads. But at the same time the advantage is lost of the compensating effect of that self-inductance in a certain frequency range with regard to the positive feedback of the anode upon the control grid belonging to it, and in the absence of such compensation the valve might tend to oscillate. This tendency to oscillate is counteracted in the QQE 06/40 by introducing two small neutralizing capacitors. Each of these capacitors is formed by the lead of one anode and a short length of wire welded onto one of the extended support rods of the control grid belonging to the other anode (see fig. 3). The capacitance is practically equal to that between an anode and its corresponding control grid. In this way a neutralization is obtained which is entirely independent of the frequency at which the valve is working.

The anodes are molybdenum plates coated on both sides with zirconium powder to reduce the secondary emission coefficient and to improve radiation of heat.

On either side of the screen grid is a U-shaped plate, called the beam-plate, which is connected to the cathode, the object of this being to prevent deflection of the electrons from the shortest trajectory. Thus these plates assist in concentrating such a space charge between the screen grid and the anodes that the secondary

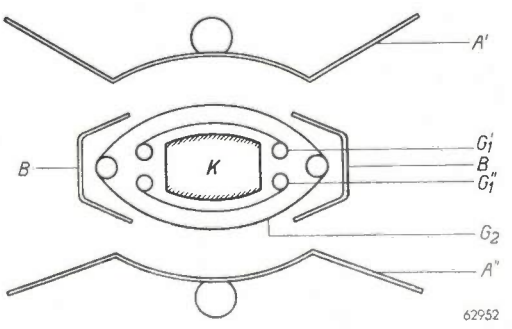


Fig. 4. Horizontal cross section of the QQE 06/40 valve. For the meaning of the letters see fig. 3.

electrons cannot reach the screen grid when the anode current is large.

Since the beam-plates prevent them from following long trajectories, all the electrons have about the same and the shortest possible transit time. Without such a measure there would be differences in transit time and at very high frequencies these differences would adversely affect the efficiency of the valve.

As is the case with receiving valves, in the QQE 06/40 a mica disc is used for fixing the mutual positions of the electrodes. This plate is screened from the strong electric field of the anodes, so that there are practically no dielectric losses in the mica, which again makes for good efficiency.

Except for the anodes, whose leads and supporting rods pass through the top end of the hard-glass envelope (fig. 3), the rest of the electrode system is mounted on a base of sintered glass [10]), into which seven rods of molybdenum have been fused. Three of these rods extend farther into the envelope than the others and carry the screening of the mica plate. This screening plate together with the beam-plates welded onto it form a framework, in which the cathode and the grids are fixed. Thus an exceptionally rugged construction is obtained, which makes the valve resistant to severe shocks.

Electrical properties

As already mentioned, the cathode is heated by means of two filaments interconnected at one end. These filaments can be connected either in parallel or in series as required, in view of the fact that some motorcars have 6 V batteries while others have batteries of 12 V; the total consumption is thus 6.3 V, 1.8 A or 12.6 V, 0.9 A respectively.

The D.C. anode voltage is max. 600 V at frequencies below 250 Mc/s, max. 400 V at frequencies above 300 Mc/s and max. 500 V in the intermediate frequency range; the screen-grid voltage is 250 V. These voltages can be derived from a rotary converter or from a transformer working together with a vibrator. The dissipation of each of the anodes may amount to 20 W and that of the screen grid to 7 W.

The input capacitance measured between the two control grids is about 6.7 pF, while the output capacitance between the two anodes is about 2.1 pF.

The feed-back of each anode upon its corresponding control grid is quite insignificant, thanks to the built-in neutralizing capacitors, so that the QQE 06/40 cannot oscillate unless feedback is purposely applied externally. In amplifiers the absence of internal feedback ensures a high degree of stability Owing to the self-inductance and the resistance of the cathode lead being extremely small, only a small driving power is needed, which can be taken, for instance, from an EL 41 valve.

The highest frequency at which the QQE 06/40 can still operate with a reasonable efficiency is about 430 Mc/s (wavelength 0.70 m). It can work at still higher frequencies but then the power gain is no greater than that of an equivalent triode.

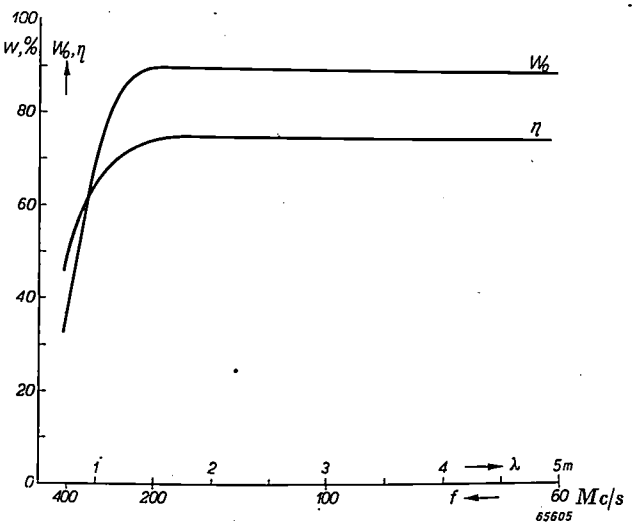


Fig. 5. Output W_0 and efficiency η [11]) of the QQE 06/40 valve as functions of the wavelength λ and the frequency f.

In *fig. 5* the output and the efficiency [11]) have been plotted as functions of the wavelength. It is seen, for instance, that at frequencies below 200 Mc/s 90 W can be generated with an efficiency of about 75%, and that at a frequency of 300 Mc/s these figures are 70 W and 65% respectively.

The double tetrode QQC 04/15

Construction

In cases where a lower output suffices there is need of a smaller and less expensive valve, and it is with a view to meeting this need that the QQC 04/15 valve has been developed, an illustration of which is given in *fig. 6*.

The construction is analogous to that of the QQE 06/40 in that the QQC 04/15 is likewise a double tetrode with one screen grid common to both the electrode systems (see the cross section in *fig. 7*).

[10]) E. G. Dorgelo, Sintered glass, Philips Techn. Rev. 8, 2-7, 1946.

[11]) As usual, here efficiency is understood to be the ratio of the output W_0 to the D.C. power fed to the anodes.

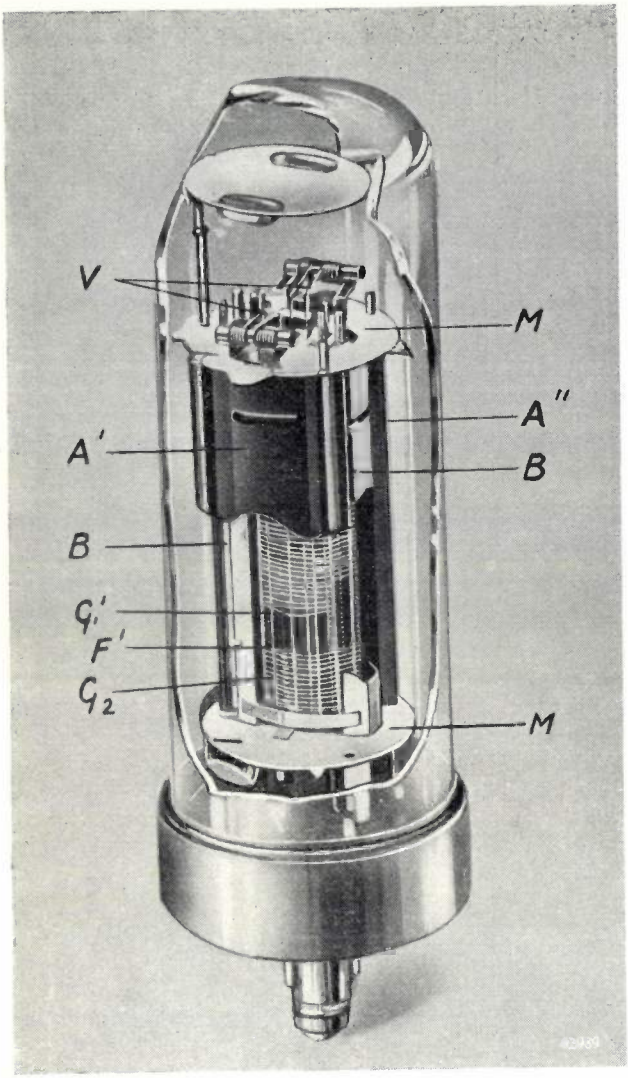


Fig. 6. Photograph of the double tetrode QQC 04/15 cut open to show the inside. F' one of the directly heated cathodes, G_1' one of the control grids, G_2 screen grid, A', A'' anodes, B beam-plates, M mica disc, V springs keeping the filaments stretched.

One point of difference, however, lies in the cathodes, which in this construction are directly heated and each consists of a V-shaped, oxide-coated filament. Such a cathode requires less heating power than that needed for a corresponding indirectly-heated cathode, whilst also the thermal inertia is much less. The cathode of the QQC 04/15, which consumes 4.3 W, reaches its working temperature 1.5 seconds after switching on, so that, in order to avoid unnecesary draining of the battery, the filament current can quite well be switched off while the transmitter is not working.

The two V-shaped filaments are connected in series and the common point is connected to a base pin.

A directly heated cathode causes a greater input damping than an indirectly heated one, and for this reason it was not necessary to use neutralizing capacitors in the QQC 04/15, but on the other hand a relatively larger driving power is needed.

The envelope is made of soft glass. This, it is true, cannot withstand such a high temperature as the hard glass of the QQE 06/40, but it has the advantage that the valve can be manufactured on the machines equipped for the mass production of receiving valves. Thus the QQC 04/15 has the appearance of a receiving valve (fig. 6), namely that of one of the key type.

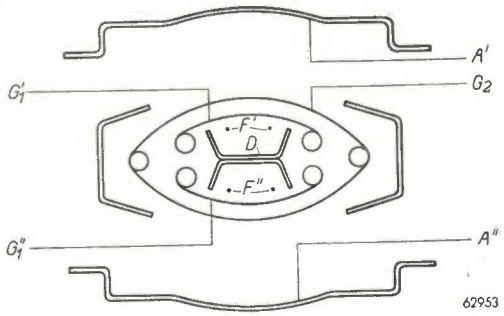


Fig. 7. Horizontal cross section of the QQC 04/15 valve. F', F'' cross section of the two V-shaped, directly heated cathodes, between which is a screen D. The other letters are as indicated in fig. 4.

A difference compared with normal receiving valves lies in the base pins, which are of chrome iron and coated over their entire length with a thin layer of copper, in such a way, of course, that the leads are vacuum-tight. The resistance of this layer at very high frequencies is much less than that of non-coated pins [12]).

The QQC 04/15 is well able to withstand the mechanical shocks occurring in automobiles and trains, just as well as the other valves used in the transmitter and in the receiver.

Electrical properties

At a voltage of 6.3 V the filament current of the QQC 04/15 is 0.68 A. When a 12 V battery is used two of these valves can be connected in series.

The D.C. anode voltage is max. 400 V and the screen-grid voltage 200 V. Each of the anodes has a dissipation of 8 W and the screen grid 7 W.

When used in push-pull the QQC 04/15 has an input capacitance of 5.7 pF and an output capacitance of 1.7 pF. The capacitance between an anode and its corresponding control grid is 0.05 pF.

[12]) Cf. K. Rodenhuis, Two triodes for reception of decimetric waves, Philips Techn. Rev. 11, 79-89, 1949 (No. 3), in particular p. 81.

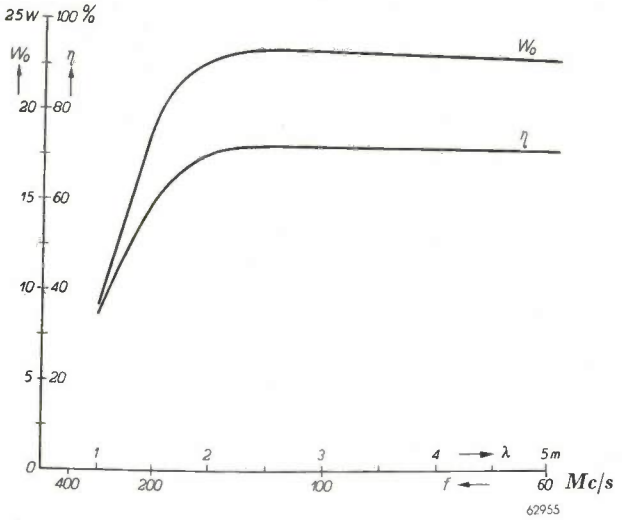


Fig. 8. Output W_0 and efficiency η [11]) of the QQC 04/15 valve plotted as functions of the wavelength λ and the frequency f.

Fig. 8 shows the output and efficiency of this valve as functions of the wavelength. The output at frequencies of 150 and 300 Mc/s, for instance, is respectively 22.5 W and 9 W, with efficiencies of over 70% and 34%.

Mobile installations with the new valves

Philips' Telecommunication Industry (Hilversum, Holland) is turning out a mobile installation, type SRR 192, with the transmitter and the receiver built together in a metal case that can be carried in the boot of a car (*fig. 9*). The control panel, the microphone and the loudspeaker are mounted in front of the driver's seat. The whole of the installation is tropic-proof.

Fig. 9. Mobile transmitter and receiver, type SRR 192, in the boot of a motorcar.

The transmitter has a QQE 06/40 as output valve and offers the choice of two frequencies within a range of 300 kc/s. Frequency-multiplication is obtained by means of three EF 42 valves and one EL 41 valve. Frequency modulation is applied.

When this apparatus is used for a fixed station it can be supplied from the A.C. mains. Between two of these sets, one mobile and the other installed at a fixed point with an aerial 25 m above the ground, in open country a range of 20 to 25 km can be covered.

For further particulars reference is made to the article quoted in footnote [3]).

One of the authors of the present article (P.Z.) has designed a smaller mobile installation which likewise works on the frequency-modulation system and has a carrier frequency of 186.24 Mc/s [13]). This employs four QQC 04/15 valves, one of which in the output stage and three for frequency multiplication.

[13]) This design is fully described in "QQC 04/15 Double Tetrode for Mobile Transmitting Equipment", a technical publication issued by the Electronic Tube Division of Philips, Eindhoven, Holland.

Summary. Following upon some introductory remarks concerning mobile radio stations such as are now being used for communication between automobiles or trains (mutually and with a fixed point), a description is given of two transmitting valves that have been specially developed for this purpose. These are both double tetrodes (types QQE 06/40 and QQC 04/15) in which the two electrode systems are so constructed that the screen grids form mechanically one whole, by this means completely avoiding the complications arising from stray self-inductance in separate screen-grid leads. In an output stage the two electrode systems of a valve are preferably used in push-pull. For frequency multiplication they can be connected in cascade. The QQE 06/40 type has an indirectly-heated cathode of a special construction, owing to which only a very little driving power and thus few stages are needed. Two built-in neutralizing capacitors provide for good stability. The cathodes of the QQC 04/15 type of valve are directly heated and consequently there is little thermal inertia. At frequencies of 200 Mc/s and below the QQE 06/40 can generate 90 W with an efficiency of about 75%, and at 300 Mc/s the output is 70 W with 65% efficiency. For the QQC 04/15 these figures are respectively about 22.5 W with over 70% and 9 W with 34% efficiency.

THE MANUFACTURE OF QUARTZ OSCILLATOR-PLATES

III. LAPPING AND FINAL FREQUENCY ADJUSTMENT OF THE BLANKS

by W. PARRISH *).

549.514.51:621.396.611.21:621.923

A manual sweeping lap motion is replaced by the cycloid motion of a machine. The mechanical abrasion by the grains of an abrasive material is supplemented by the molecular corrosion of an etching solution. These are the principles which have made it possible to increase the production of quartz oscillator-plates a hundred or a thousand times and to give the plates permanently the desired properties.

A quartz plate intended for the stabilization of the frequency of a short-wave transmitter working on a wavelength between about 3 m and 150 m is usually cut from the crystal in the orientation of the so-called AT or BT cut, and is made to vibrate

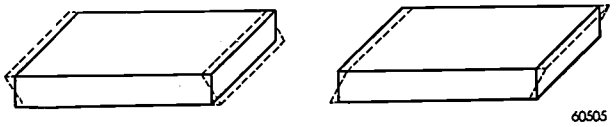


Fig. 1. Oscillator-plates according to an AT or BT cut are square, with sides from 3 to 50 mm long (normally about 12 mm). These plates are made to vibrate in a thickness shear mode. Two opposite phases of the deformation are shown by dotted lines.

in a thickness-shear mode (*fig. 1*). The characteristic frequency of plates vibrating in this way depends principally upon their thickness and is inversely proportional to it. The relation between frequency f and thickness D ($f \times D$ = constant = K) is represented graphically in *fig. 2* for the two cuts mentioned. In practice, plates are used with a thickness varying from about 0.8 mm (AT plate with a frequency of 2 Mc/s) to about 0.3 mm (BT plate with a frequency of 9 Mc/s). For transmitters with frequencies higher than 9 Mc/s thinner plates are not usually used, but a thick plate is allowed to vibrate with an odd harmonic of the mode of vibration illustrated, or oscillator circuits with frequency multiplication are used [1]).

Since it follows from the relation mentioned that $df/dD = -K/D^2$, it is obvious that in the case of thin plates a slight change in thickness causes a large change in resonance frequency. In the case of a BT-cut crystal with $f = 9$ Mc/s a change in thickness of for instance 10^{-5} mm corresponds to a frequency change of 320 c/s. A transmitter which is stabilized by the 11th harmonic of this crystal, with the deviation in thickness mentioned, would operate more than 3 kc/s away from its nominal frequency (not exactly $11 \times 0.320 = 3.52$ kc/s, since the "harmonics" of a vibrating plate are not in exact integer ratios to each other). In order to be able to use all the plates produced immediately for definite channels in the intended wavelength region (the channels are usually about 100 kc/sec wide at the highest frequencies considered), the required frequency is given with a tolerance of for example 500 c/s. The thinnest blanks must then be finished to have the necessary thickness within an accuracy of about 2×10^{-6} mm, in order to be suited for use in the 11th harmonic.

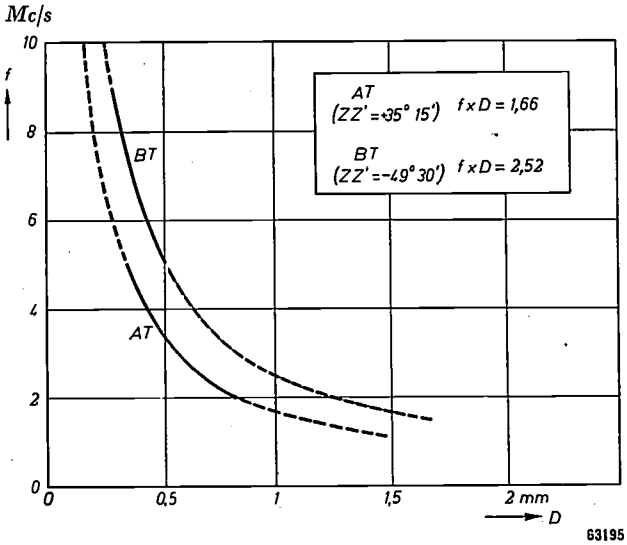


Fig. 2. Relation between resonance frequency f and thickness D of quartz plates according to the AT and BT cut.

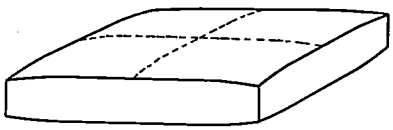
*) Philips Laboratories, Inc., Irvington-on-Hudson, N.Y., U.S.A.

1) The lower limit of 0.3 mm for the thickness was imposed by the requirement of reliability in mobile equipment for wartime applications, which could be exposed to severe shocks. To-day's mobile transmitters are equipped also with thinner plates, down to about 0.15 mm. Lapping of blanks is possible to even smaller thicknesses, of about 0.06 and even 0.03 mm. — For the position of the AT, BT and other cuts in the crystal as well as for the manner in which a plate can be cut in the desired direction see two preceding articles in this periodical: W. Parrish, The manufacture of quartz oscillator-plates, I. How the required cuts are obtained, Philips Techn. Rev. 11, 323-332, 1949 (No. 11), and II. Control of the cutting angles by X-ray diffraction, Philips Techn. Rev. 11, 351-360, 1949 (No. 12).

In practice this is done in the following way. The blanks are sawed out of the crystal with a thickness of 1 to 1.2 mm and then lapped to the required thickness of 0.3 to 0.8 mm. The fact that so much must actually be lapped off — or rather that it is necessary to begin with blanks so much thicker than the finished plates — will become clear later.

Since the resonance frequency is so sensitive to a change in thickness and because frequencies can easily be measured with very great accuracy, the process of lapping a quartz plate is not checked by a precision measurement of the thickness reached, but directly by a determination of the resonance frequency obtained. Therefore, in the art one speaks of "lapping off a certain number of cycles" — which means that the thickness is lapped off until the frequency has risen by that number of c/s.

In order to obtain a crystal of good activity, i.e. one which upon weak excitation vibrates with a reasonably large amplitude and which dissipates little energy, the surface must be smooth and completely without flaws and the upper and lower surfaces of the crystal must be lapped so that they are exactly parallel. At the edges, however, and especially at the corners the thickness must be one to several microns less than in the center part; see the much exaggerated sketch in *fig. 3*. This shape has been found empirically to be favorable for the avoidance of the simultaneous occurrence of undesired modes of vibration of the plate, which would have an unfavorable effect on its activity.



63196

Fig. 3. "Contour" of quartz plates required for high "activity". The deviations from plane parallelism are drawn very much exaggerated.

The lapping of a blank can be done by hand, by pressing the blank gently against a glass surface covered with abrasive and moving it with a sweeping motion over the glass, preferably in the form of figure eights (8). The desired contour can also be given to the blank by distributing the pressure in a suitable way during the lapping. The blank is first lapped for a time with two fingers pressing on diagonally opposite corners and then for a time with two fingers on the other two corners. Care must at the same time be taken that the blank travels uniformly over all parts of the glass surface so that the latter, which is of course also somewhat lapped off, may remain perfectly flat.

For the mass production of quartz plates such as had to be organised in 1942 in the United States at very short notice because of the necessities of the war, lapping by hand was naturally very undesirable and thus mechanical methods of lapping were sought. A solution which, as far as speed and precision obtained are concerned, gives excellent results was found in the planetary lap machine. This machine was developed by G. C. Hunt and P. R. Hoffman [2]). The lapping procedures with this machine have been critically investigated and further developed by the North American Philips Co. in a pilot plant set up for that purpose, and they were being used at the end of the war in numerous quartz platc industries [3]).

The planetary lap machine

Construction

The principle of the planetary lap machine will be discussed with reference to *fig. 4*. On a stationary circular iron lap plate 30 or more blanks to be lapped are laid in work holders of gear-wheel form as shown in the figure; the square blanks lie loosely in the pentagonal holes of the holders. By means of an inner gear turning in a central bore of the plate and an outer ring gear rotating around the circumference of the plate the work holders are brought into a complex rotating motion around their own centres and around the centre of the lap plate (hence the name of the machine; cf. also the planetary driving mechanism which was used for example in some of the first steam engines). Each blank describes a broad sweeping lap motion over the plate, which is equivalent to the motion in hand lapping.

The work holders are slightly thinner than the quartz plates to be lapped. A second lap plate is laid on top of the arrangement, so that it rests upon the upper side of the blanks, and is held loosely in position by a slotted arm (it thus remains stationary; see *fig. 5*). A suspension of the abrasive — which will be dealt with later in detail — in a lubricant is introduced between the two lap plates via feed holes in the top plate. The abrasive accumulates in the serrations cut into the lapping

---

[2]) G. C. Hunt, U.S. Patent No. 2,314,787, March 23, 1943; P. R. Hoffman, U.S. Patent No. 2,308,512, June 19, 1943.

[3]) A survey is given in W. Parrish, Machine lapping of quartz oscillator-plates, American Mineralogist **30**, 389-415, 1945. See also: W. L. Bond, Chap. IX in R. A. Heising, Quartz crystals for electrical circuits, D. van Nostrand, New York 1946.

surfaces of the two plates. Sweeping over the serrations all 30 blanks are lapped simultaneously and uniformly, upper and lower surfaces alike.

Closer examination of the movement of the blanks being lapped

It will immediately be clear to the reader that an enormous production can be obtained with such a machine, and also that the requirement of parallelism of the two sides of the crystal can easily be met in this way. It will be less obvious however that with the lap machine alone, without further treatment, the typical contour shown in fig. 3 can be obtained. In order to understand this we must consider the movement of the blanks in the lapping process more closely. Apart from this, it is in any case worth while to examine this peculiar motion.

Let us first consider the case in which the inner and outer gears rotate in the same direction at the same number of revolutions per unit time. In that case the work holders will move as if they were rigidly connected to inner and outer gear. This amounts to the same thing as if the work holders were stationary and the two lap plates slid above and below them. In various respects the result of this motion will not correspond to what is desired.

In the first place the outer crystals cover a longer distance in the abrasive than the inner ones and are thus lapped thinner. It might seem that this would not be objectionable: it would mean one lap would produce a series of crystals of different thicknesses which would require sorting later as to thickness for the different channels (frequency regions). Although sorting and classification is

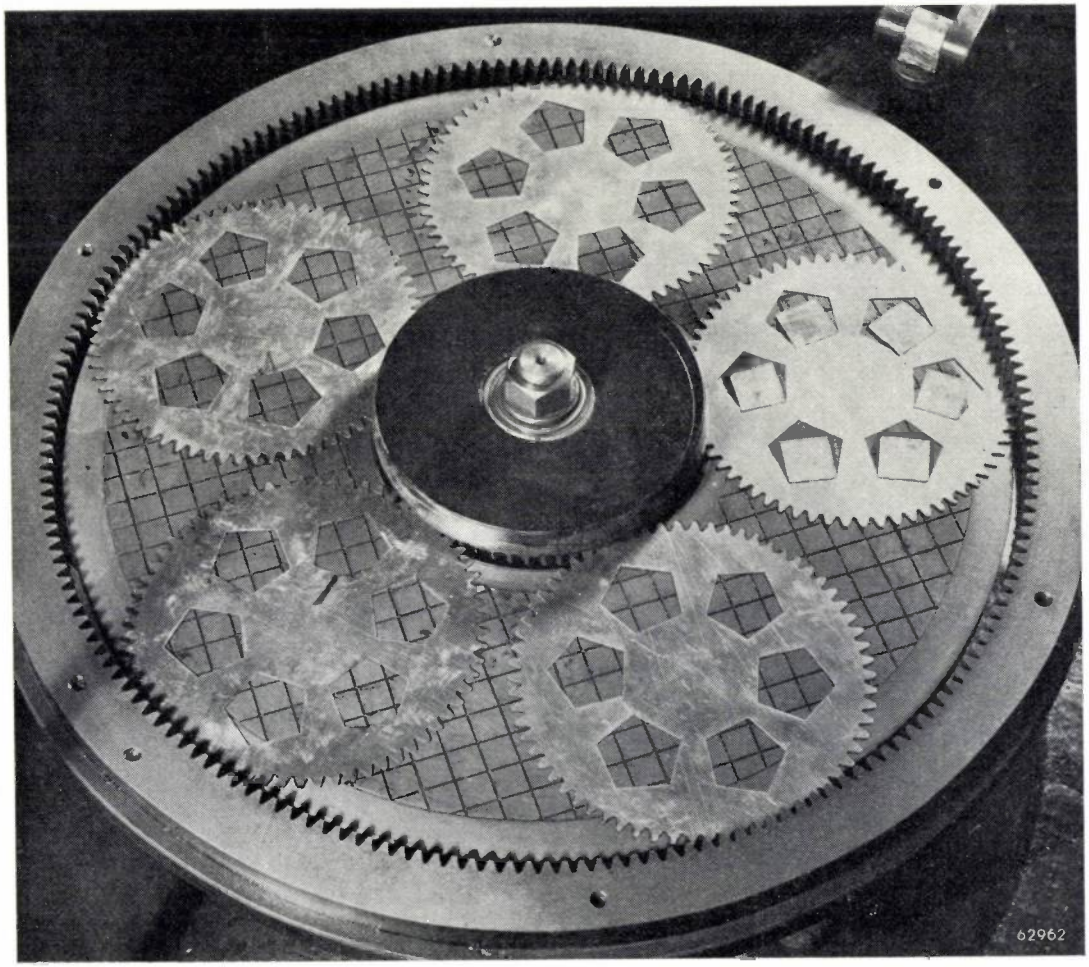


Fig. 4. The planetary lap machine. The serrated lap plate is stationary, the outer ring gear and inner gear rotate. The five gear-shaped work holders, each of which can hold six quartz blanks for lapping in the pentagonal holes (only one of the holders is filled) describe a complex motion over the plate. A second lap plate is laid on top of the arrangement so that the thirty blanks are lapped on both sides simultaneously. For smaller oscillator-plates holders with 8 and 11 holes are used, so that a lap then consists of 40 or 55 plates, respectively.

possible, such a method would be irrational. It would be necessary to keep large stores of quartz crystals on hand and this would mean much administration, etc. For smooth production it is on the contrary much better to be able to make all the crystals of a single lap for a given channel, which means that the frequency spread of the 30 or more crystals at the end of the lap should not be more than 10 to 15 kc/s [4]).

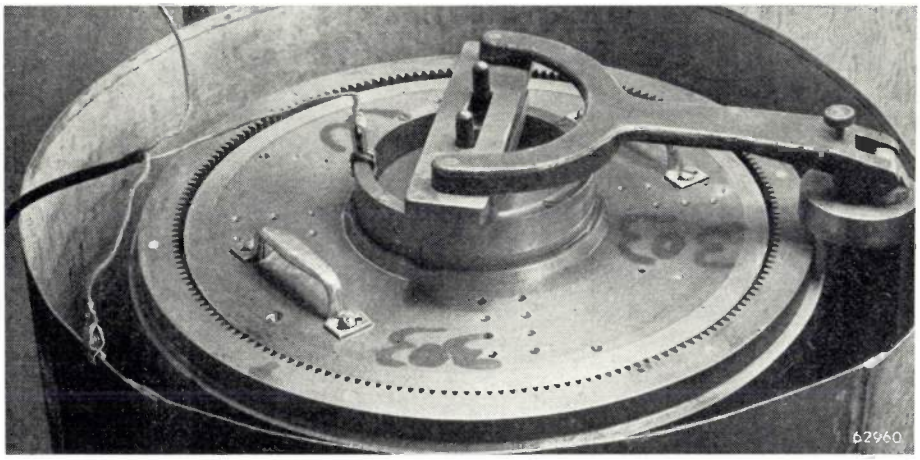


Fig. 5. Planetary lap machine in use. On top of the lap plate with work holders as shown in fig. 4 the second lap plate is laid and floats upon the crystals. The slotted arm prevents this plate from turning with the holders. The abrasive suspended in a lubricant is fed through the holes in the upper plate and flows between the two lap plates. The outer metal shield catches the spattering abrasive suspension. The significance of the electrical connections is explained below.

In the second place — and this is worse than a mere inconvenience — the simple circular motion would by no means lead to the desired contour of the crystals. The upper and lower surfaces would not be parallel, but all the crystals would be wedge-shaped: in rotating, the same edge of the blanks would always come into contact with the fresh abrasive and their surfaces would therefore be lapped off more on that edge than on the opposite edge.

In the third place the iron lap plates would not long remain flat, since the stronger lap action near the circumference would naturally also affect the lap plates.

Let us now compare this with the actual motion of the blanks. The inner gear is made to rotate faster than the outer ring gear, about 3 times the number of revolutions (in the same direction). The result is that each work holder with its six blanks in addition to the "translation" over the surface of the lower lap plate, also executes a rotation about its own centre. The resulting relative motion can for our purposes best be represented as if the holders, rotating around their own centres, remain stationary while the two lap plates slide along them above and below. In this motion each blank travels periodically back and forth between the inner and outer edge of the lap plates. All the blanks and also the whole surface of the lap plates are now lapped to practically the same extent (we shall consider the statistical fluctuations in this below).

A further feature of the actual motion is that no wedge-shaped crystals will be produced. In fact, due to the rotation of the work-holder each of the four edges of a blank gets its turn in receiving the fresh abrasive. Moreover, each blank also "spins" around its own centre, in the pentagonal hole of the work holder, as is explained in *fig. 6* and the accompanying legend; hence, the "turns" of the four edges are not correlated with invariable positions of the work holder, but, on the contrary, each edge gets its turn successively in all possible positions of the holder (with the corresponding, more or less widely differing relative velocities of the blank and the lap plates!). Thus, all four edges of the blank are assured of a completely identical treatment and fine plane-parallel crystal plates are obtained.

As regards the special contour of the oscillator-plates, owing to the motion described the blanks automatically will acquire a form similar to that

[4]) For BT cuts of 6-8 Mc/s. This tolerance is much larger than that mentioned in the beginning, because a final frequency adjustment follows the mechanical lapping. We shall return to this in the last section.

drawn in fig. 3. In fact the lapping action is stronger at the edges — which receive fresh abrasive — than in the centre, and strongest at the corners (which belong to two edges). Therefore, the blanks at the edges and especially at the corners are lapped thinner than in the centre, and the difference primarily will even be greater than desired! Thus, the real problem is to restrict the lapping of edges and corners. This is done by a very simple device: the ring-shaped lap surface of the two iron lap plates is made slightly narrower than the ring covered by the blanks in their cycloid motion. Thus at the inner and outer edges the crystals slightly override the periphery of the lap plates and it should be noted that, due to the spinning motion of the crystals, at successive times that the crystal arrives at the periphery different corners of it will override the latter. On an average the lapping action at the edges and especially at the corners is thereby slightly decreased, equally for all four corners, and by a suitable choice of the amount of override the crystals can be given with good approximation the contour which is desired for a good activity.

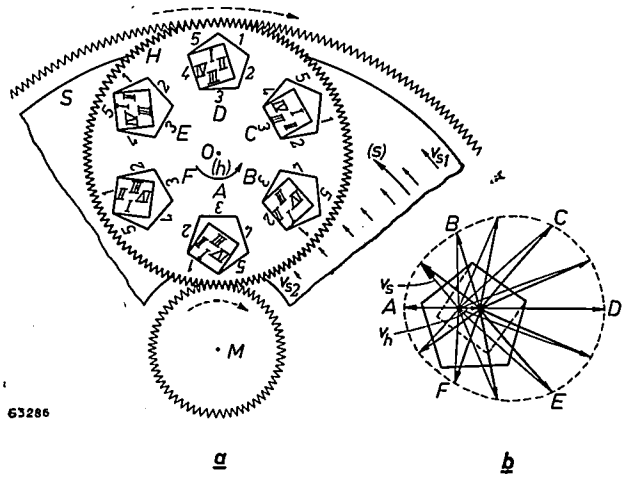


Fig. 6. The relative motion described by a work holder H with respect to the lap plate S can be built up of two components: the work holder revolves about its own stationary centre O (motion h); the lap plate rotates about the centre M of inner and outer gears (motion s). Each crystal passes successively through the positions A-F. Moreover, each blank in the pentagonal holes in the work holders can execute a spinning motion around its own centre by tipping with successive edges against successive sides of the pentagon. This tipping motion comes about in the following way: If only the centrifugal force of the motion h acted on the crystal it would always remain with side I against side 1 of the pentagon. However, the frictional force k_{hs} is also exerted on the crystal by the lap plates and this is opposite in direction to the resultant absolute velocity of the crystal (composed of the velocities v_h and v_s); this force is so much larger than the centrifugal force that the latter in this case may be neglected. Passing through the positions A-F by the crystal amounts to a rotation of the direction of v_s with simultaneous variation of its magnitude between v_{s1} and v_{s2}, while v_h does not change in magnitude or direction (with respect to the pentagon). In fig. b the resulting force k_{hs} on the crystal is shown for a number of positions for a whole revolution of the work holder. As soon as this resultant force in rotating passes a corner of the crystal, the crystal is tipped against the following side of the pentagon.

The lap plates and the work holders

The two lap plates are made of normalized meehanite. Experiments have also been carried out with lap plates of hardened tool steel, but the advantage of the greater hardness of the steel which should permit a higher lap speed is cancelled by the fact that it resists the embedding of abrasive particles; the latter process ("charging" of the plates) readily occurs with the meehanite and greatly enhances its lapping action. Moreover, steel has the drawback that it is not so easy to make the necessary serrations in it. (The photograph on page 177 in this issue shows how this is done in the case of the meehanite plates.) Smooth steel plates do not lap faster than the meehanite plates with serrations. Glass lap plates give a satisfactory speed of lapping but have the disadvantage of quickly getting out of flat by the abrasive action and that the blank and also the work holders tend to stick to the upper lap plate when it is lifted off.

The pressure on the blanks required for lapping is obtained by giving the upper lap plate a suitable weight. This is not critical. According to the thickness of the blank and the abrasive used weights between 15 and 28 lbs are chosen, so that the average weight per crystal is about 1/2 to 1 lb.

For a uniform distribution of the pressure it is important that the lap plate should rest freely on the blanks, i.e. that the arm which prevents the plate from rotating should not transmit to it any unsymmetrical vertically directed forces. Fig. 5 shows how this is accomplished.

The two lap plates, ground flat with a large surface grinder, are lapped-in by placing them in the machine and allowing the machine to work for a time with the work holders replaced by three massive meehanite gears. The same abrasive is used as will later be used for lapping the quartz. The coarse scratches on the lap plates caused by the previous working with a coarse-grained grinding disc are thus replaced by myriads of nearly invisible grinding scratches which correspond to the grain size of the fine abrasive. The quality of the surface of the lap plates then no longer changes from one lap of quartz crystals to the next. Moreover, during the lapping-in the meehanite plates already get "charged" to some extent (see above).

The work holders are made of linen-base "Bakelite" or "Philite", or of zinc. The first material has the

disadvantage that it tends to become warped, but compared with zinc it is less apt to lead to chipping of the edges of the blanks as they are tipped in the pentagonal holes, it causes less wear on the teeth of the outer and inner gears and it causes less damage when a crack-up occurs while the machine is running. The holders must be at least 0.05 mm thinner than the final lapped thickness of the quartz plates. The five work holders used at one time may not differ from each other in thickness by more than several hundredths of a mm. Well finished work holders can be used several times before they are worn out.

The material of the work holders is much softer than iron, but their teeth are continually "lubricated" with the splashings of the suspension of abrasive so that the already noted wear of inner and outer gears — especially the inner — is quite noticeable: at the level on the gears where the teeth of the work holders lock, the gears are slotted on one side of the teeth. After a number of laps the lower lap plate must be lowered slightly between inner and outer gears so that the teeth of the work holders lock into the gears at a slightly lower level. This operation is made easy by the fact that the lower side of the bottom lap plate has three feet resting on three holders and in addition to these three feet three series of 14 successively shorter feet in a step-like arrangement around the periphery, so that when the plate is lifted and slightly rotated it falls into a slightly lower position. There are thus 15 positions which can be used successively. Then the inner and outer gears can be turned upside down in order to bring into play the other side of their teeth, which has not yet been worn; again all 15 positions can then be used. Only then is it necessary to replace the inner and outer gears by new ones.

Control of the stopping point of a lap

Since lapping with the planetary lap machine proceeds quite rapidly (one lap takes only a few minutes) and since the tolerances in thickness of the quartz plates are so small, it is very important that the lap be stopped at the right moment. The rate of lapping depends upon very many factors, and experience has shown that it is impossible to take these factors into account in such a way that it would be possible to determine the required lap time in advance with sufficient accuracy to meet the close tolerances. It would, therefore, seem obvious to use the same method as that used in the lapping of separate crystals by hand: stop the lapping as soon as it may be assumed that the correct frequency has practically been reached, determine the resonance frequency of one (or more) of the crystals in an oscillator, lap cautiously further according to the difference between the frequency measured and the required frequency. This process can then be repeated several times until one has approached sufficiently close to the desired frequency. In machine lapping however this method is not practical. The measuring of the resonance frequency of a crystal does not by itself offer difficulties — it takes only a few minutes — but the opening of the machine and removing of a crystal is troublesome, time-consuming and always adds the additional risk of a crack-up. It soon develops that a disproportionate amount of time and work is spent on control, as compared with the actual lapping.

An elegant solution of the problem has become possible due to the surprising observation that the resonance frequency of the crystals can be measured while the machine is running. During the lapping the crystals receive tiny mechanical shocks from the grains of the abrasive. These shocks cause the AT and the BT plates to vibrate each time, and this they do in the "right" mode with the corresponding characteristic frequency. The result of this vibration is the occurrence of opposite electric charges on the upper and lower surfaces of the quartz plate (normal piezoelectric effect), which change signs in the rhythm of the mechanical vibration. Hence a weak alternating voltage occurs on the two lap plates (which with the crystals between them act as a condenser) and this voltage has a frequency equal to the resonance frequency of the quartz plates (or rather: equal to a kind of average resonance frequency; this point will be reconsidered below). If care is taken that the two lap plates are not short-circuited somewhere outside the crystals, the alternating voltage can be picked up with a sensitive standard short-wave receiver and its frequency measured. The latter is done by calibrating the tuning dial of the receiver (at the frequency the lap is to be stopped) with a calibrated variable oscillator which itself may be calibrated with a standard quartz crystal.

During the lapping the operator continually moves the tuning knob back and forth slightly in order to keep the signal that he hears in head phones at maximum intensity. The position of the tuning knob is thus shifted gradually toward higher frequencies as the plates are lapped thinner. When the tuning has reached the previously indicated calibrated piont, the operator stops the machine with a pedal switch. (In principle it is possible to make the control and stopping automatic.)

Fig. 7 shows the set-up for this method. In fig. 5 the necessary electrical connections to the lap machine may be seen. The connections must be in shielded cable in order to minimize electrical interference, because the signal of the crystals has an

intensity of only a few microvolts and could thus easily be drowned out by interference or (with inefficient coupling) by noise.

For the purposes of this method work holders of "Bakelite" ("Philite") will be preferred above those of zinc. Although the work holders do not make contact with the upper lap plate, the zinc holders cause a considerable increase in the capacitance between the plates and thus a smaller signal-to-noise ratio.

Fig. 7. Set-up of a planetary lap machine with apparatus for controlling the stopping point of the lap.

The frequency spread

When at the end of one lap the resonance frequency of all 30 crystals in the 5 work holders is measured, precisely the same frequency is by no means found for all of them. The frequencies are found to be statistically distributed about a certain mean value. In view of the fact already mentioned that it is preferable to be able to use all the crystals of one lap for the same communication channel, the frequency spread was a serious obstacle at the beginning of mass production. In order to explain the remedy which was found later we must consider the lap process somewhat more closely.

The surface of a quartz plate becomes smoother the finer the grain of the abrasive used. It has been found that one may be sure of the smoothness required for good activity when one laps with Al_2O_3 303 i.e. a powder of Al_2O_3 (optical flour) with a grain size of about 15 microns. However, the speed of lapping with such a powder is very low, especially when most of the coarse inequalities of previous sawing or grinding operations have been removed. Impossibly long lap times would thus be required.

The method generally followed therefore is that a lot of, for instance, thirty crystals is lapped in at least two and preferably three lapping stages. A coarse powder is used at first, which removes the deep sawing marks present on both surfaces of the crystal plate and thus the largest part of the thickness which must be taken away. The process is repeated with a slightly finer powder which quickly removes the finer scratches and inequalities

left by the first powder, and hence does not need to remove much of the thickness. And one finishes with the finest powder mentioned which laps off the inequalities of the second powder and gives the plate the desired smoothness.

According to our experience the following is a very efficient program:

1) Lapping with SiC 320, i.e. silicon carbide with a grain size of 31.5 μ, to the final thickness + 0.27 mm. This will take about 4 minutes.

2) Lapping off about 0.2 mm with SiC 600 (grain size 17.5 μ), in about 6 minutes.

3) Lapping for about 12 minutes with Al_2O_3 303, thus removing about 0.07 mm.

Assuming a thickness of 0.1 to 0.2 mm to be removed from both sides of a blank in the first lapping stage, one arrives at a total thickness to be lapped off of 0.4 to 0.7 mm, as was mentioned in the introduction.

It is necessary to perform the three laps on different machines in order to eliminate any risk of contamination of the finest abrasive by traces of those previously used. In order that the machine used for the first lap should not stand idle for long times, a group of 5 or 6 machines should be used: one for stage (1), one or two for stage (2), and three for stage (3). The first machine is run with a relatively heavy upper lap plate and at a fairly high speed (outer ring gear 120-150 r.p. min.). The last three machines have the lightest upper lap plate and the lowest speed (35-60 r.p. min.).

With such a set of machines and with this method, operating 20 hours a day, 5000 to 6000 lapped quartz plates can be produced per day.

By this method, under certain circumstances, at the end of the third lapping stage of a lot of crystals a frequency spread of 60 to 125 kc/s was found among the crystals (difference between the highest and the lowest resonance frequencies occurring); in a few cases it was even 200 kc/s. This is much larger than is permissible for good production control. Only where the spread remains below the above-mentioned limit of about 15 kc/s can all the crystals of one lap be subjected together to the final treatment (see below) and need not be further sorted for different communication channels.

It is easy to understand that the spread in thickness of crystals present after the first and second lapping stages is not very objectionable. In making up the first lap one naturally chooses crystals which do not differ very much in thickness (for instance not more than 0.1 mm, which can easily be checked with a micrometer), and then because of the symmetry of the set-up the spread in thickness upon lapping cannot rise above a certain limit [5]). For an investigation of the occurrence and the nature of the spread it is therefore sufficient to consider the situation before, during and after the last lapping stage.

This has been done in an investigation in Dobbs Ferry, in which more than 10,000 quartz plates were measured repeatedly during lapping, and in which various lapping procedures were employed. The most important result was the discovery that the spread is small, namely about 10 to 20 kc/s at 8 Mc/s for the plates of one work holder, and that the large spread often occurring in one lap can be ascribed to differences between the work holders.

This discovery led to a method to diminish the spread, a method simple enough to be introduced into the factory. The third lapping stage is interrupted at a certain moment, shortly before the end of the expected lap time, all the crystals are taken out of the machine, turned upside down and then distributed in a different way over the five holders so that every work holder now contains plates of the most divergent frequencies ("transposition"). When the lapping is continued the large differences are very quickly equalized since the thickest plates in each holder are by far the most strongly lapped. The lapping may now, however, not be continued too long, since otherwise a greater spread again may develop between the holders! The moment of transposition must therefore be carefully chosen and adapted to all the circumstances of the lapping process. We found a transposition at an average thickness of the plates of 0.015 $\pm$ 0.005 mm above the desired final value to be the most suitable. If the transposition is performed somewhat later the spread within each holder will not be sufficiently closed before the prescribed final thickness has been reached.

In *fig. 8* the frequency of a lap of crystals is given at different moments for a typical case. The effect of the transposition can easily be seen. Thanks to this method one may be assured that not more than about 3-5% of the quartz plates of a lap will fall outside the permissible frequency region, 15 kc/s wide.

The frequency spread naturally leads to the fact that in the above described method of controlling the stopping point of the lap one obtains signals which cover a whole band of frequencies.

---

[5]) In fact the spread remains within the errors of actual thickness measurements feasible in a manufacturing process. Thus it would be useless to attempt by actual thickness measurements to match a group of crystals for the next lapping stages closer than the lapping process itself produces. Frequency measurements would render this possible, but they consume too much time for such grouping procedures; moreover, most of the blanks do not yet oscillate in the early lapping stages (cf. below, in connection with fig. 8).

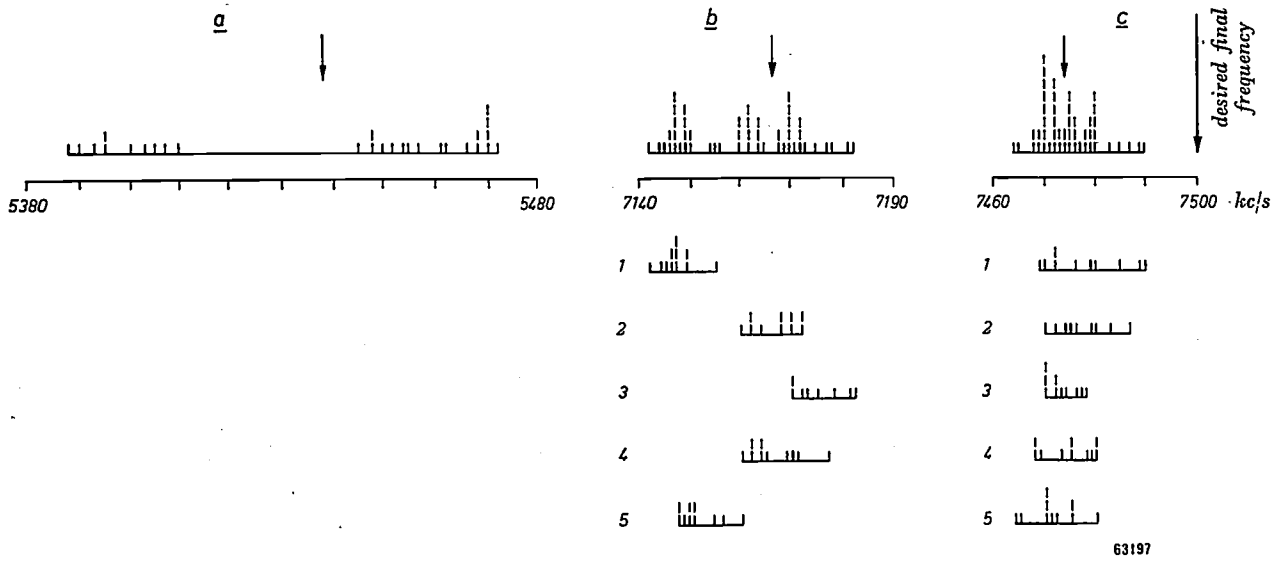


Fig. 8. The frequency spread of a lap,
a) at the end of the second lap stage;
b) after the greater part of the third stage, namely at the moment of transposition; each of the five work holders shows a relatively small spread, but their mutual spread is large;
c) after transposition and conclusion of the third stage.
Each stroke in the vertical line represents one quartz crystal. At the end of the second stage not all the crystals have sufficient activity for oscillating. Only the oscillating crystals can be measured; this will explain why there are more crystals indicated in *b* and *c* than in *a*. The fact is that after the third stage the lap often will contain less crystals than after the first and second stage, as during the lapping some crystals may be broken. These places are not filled by other crystals, for the reason pointed out in note [5]).

The intensity of the signal, however, is found to exhibit a sufficiently sharp peak at the nominal frequency to make the method applicable. The frequency distribution of the signal can be recorded with a recording meter and a good idea of the spread can thus be obtained. *Fig. 9* gives an example of such a recorded distribution curve. The frequencies of the individual crystals of the lap measured immediately after the recording are plotted under the curve.

Final frequency adjustment by etching

The quartz crystals lapped with the planetary lap machine have resonance frequencies which still lie about 10 to 25 kc/s below the desired value. The final adjustment to the desired value within the tolerance of 500 c/s (or closer depending upon the requirements) takes place separately for each crystal. In the beginning the final adjustment was performed carefully by further hand lapping (supplemented if necessary by a small single crystal lap machine). It was found, however, that after some time the crystals thus finished, which had passed all the tests, whether they were used or only kept in store, no longer met the specifications. Their resonance frequency was higher and their activity lower. The situation caused by this unexpected ageing phenomenon at a certain time in 1943 appeared to be catastrophic: millions of oscillator-plates stored in the army magazines of the U.S.A. were found to have become unusable!

An elaborate investigation [6]) led to the conclusion that the ageing was probably caused by a gradual spalling off or recrystallization of the outer layer of the quartz plate. This layer upon lapping with even the finest abrasive is subjected to a treatment which when observed on a submicroscopic scale can be called very rough; after this treatment the layer contains numerous submicroscopic crevices, hills, overhangs etc. In the course of time, and especially through the agency of adsorbed moisture, the projections weather and the effective thickness of the plate decreases, only fractions of a micron to be sure, but this already causes a considerable rise in the resonance frequency. At the same time the loose dust-like deposit of quartz particles remains on the surface and causes a certain damping which has an unfavorable effect on the activity.

The phenomenon can be rendered harmless, as was later found, by removing prior to the testing of the crystals the layer of quartz which is subject to deterioration. This is done by etching in solutions of ammonium bifluoride of a certain concentration and temperature. The effect of etching can be

[6]) Cl. Frondel, Final frequency adjustment of quartz oscillator-plates, American Mineralogist 30, 416-431, 1945.

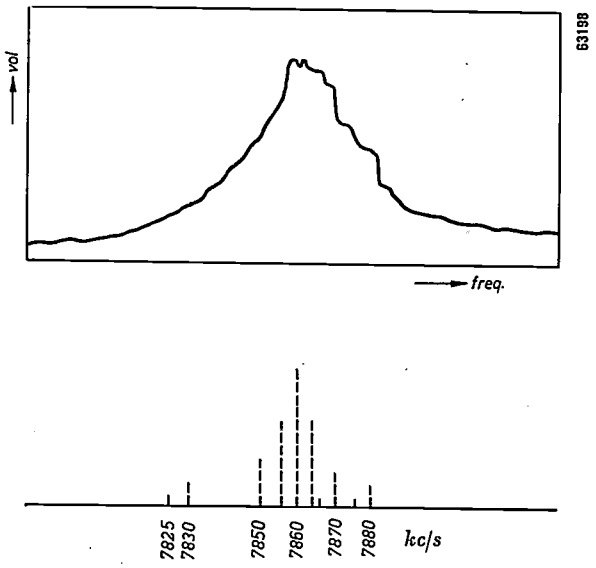


Fig. 9. *a*) Graphic recording of the intensity of the signal produced by 30 quartz crystals in a planetary lap machine as a function of the tuning frequency of the receiver.
b) Individual resonance frequencies of the quartz crystals of (*a*), measured immediately after recording the graph and stopping the machine. Each stroke represents one quartz crystal.

demonstrated by a simple experiment. 50 kc/s are lapped off from a series of crystals with a rather coarse abrasive (e.g. SiC 320) so that a severely damaged surface layer is obtained. A thickness varying between 0 and 10 kc/s is then etched off. The ageing when it occurs takes place rather quickly on the damaged surface. In *fig. 10* the observed ageing is represented graphically: the unetched plates increase by 1.5 kc/s in five days, the plates from which 10 kc/s had been etched off, after a very slight increase in frequency in the first three to four hours, showed no further ageing.

It is now prescribed in all quartz oscillator specifications in the U.S.A. that finished AT and BT plates should have a final layer of at least 1 μ thickness etched off[7]). Thanks to the limited frequency spread it is possible to subject all the crystals of one lap of the planetary lap machine to the greater part of this etching process together. In doing this, the process of course must be controlled with a view to those plates which have the smallest deviation from the final frequency. The final adjustment is by individual etching of each plate under continuous control of the resonance frequency.

[7]) In the case of BT plates for 9 Mc/s this means, as may be derived from the formulae given with fig. 1, that the frequency rises by more than 32 kc/s. In the case of a lap of quartz plates for a nominal frequency in the neighbourhood of 9 Mc/s, therefore, the end of the last stage with the planetary lap machine, expressed in kc/s, must lie considerably farther below the final frequency than in the example previously mentioned.

In this connection it should be pointed out that in lapping and etching, which amount to the removal of material, the resonance frequency is always increased, so that the desired frequency can only be approached from one direction. The possibility also exists of lowering the frequency of a given quartz plate, namely by exposing it to an intense X-radiation. Attempts have been made to put this remarkable phenomenon to practical use (see the article cited in footnote [6])). It was found, however, that the lowering of frequency produced, together with an accompanying darkening of the quartz, disappeared again after some time and particularly at elevated temperatures. This was unfortunate because the method had great advantages; not only could crystals be salvaged which had exceeded their nominal frequency up to 0.02% due to overfinishing, but the crystals could also be irradiated while they were in operation in a valve oscillator and thus it was possible to follow the approach to the desired frequency continuously on a measuring instrument.

However, the same possibilities are now obtained in a different way. Most oscillator-plates made to-day have plated electrodes. The plating, which is put on the crystal either by evaporation in a bell jar or by chemical deposition from a solution, adds to the mass of the vibrating crystal and thereby contributes in determining its resonance frequency. It is thus possible to adjust the frequency a little either down or up by depositing or removing some of the plating without touching the quartz plate itself; and in the evaporation method, similar to the irradiation procedure, it is also feasible to approach the desired frequency under continuous control as described above.

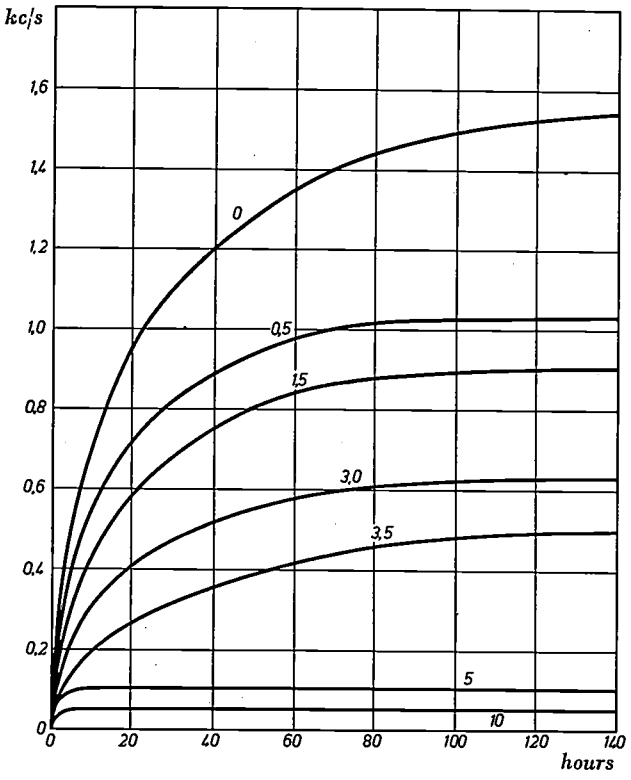


Fig. 10. Ageing observed in BT quartz crystals for about 8 Mc/s, from which 50 kc/s has been lapped off with a coarse abrasive (SiC 320) and then a layer varying from 0 to 10 kc/s etched off. The resonance frequency measured is plotted as a function of the number of hours elapsed since the etching. (From C. Frondel, Amer. Mineralogist **30**, p. 422, 1945.)

Plates finished by etching are also not perfectly stable as to the frequency. Where extreme requirements are made as to stability, thus where every trace of ageing must be avoided, the best results are apparently obtained, according to recent investigations, by heating the quartz plate to 500 °C, i.e. just below the transition point from α to β quartz, and then cooling it slowly [8]).

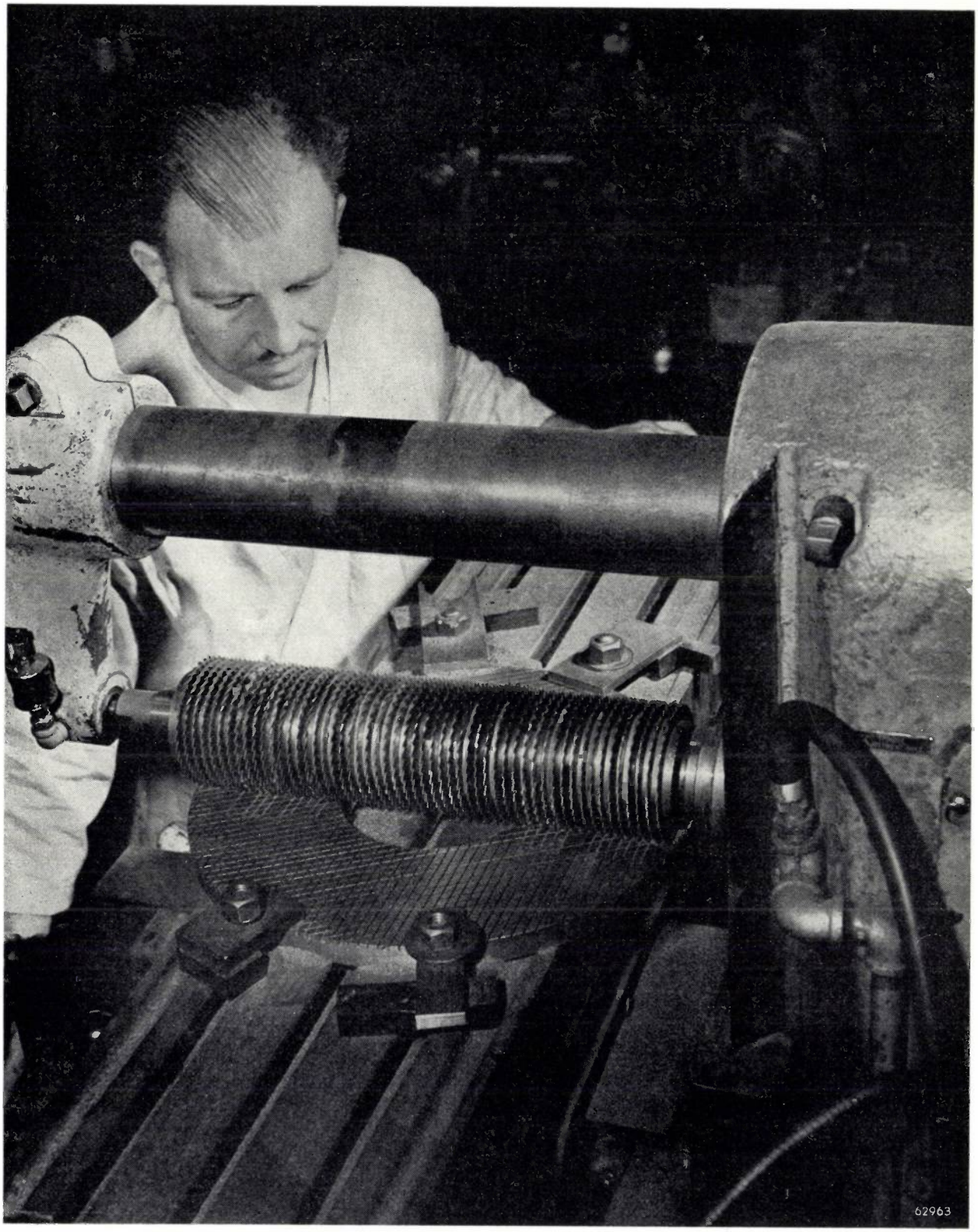
This concludes the survey given in three articles of the development of the manufacture of quartz oscillator-plates in the war years. In general the methods followed by the quartz producers still working are similar to those here described. The number of plates produced is of course very much smaller. However, the fact that such efficient methods of manufacture are available cannot help but have a stimulating effect on the use of quartz plates, so that the applications in peace time of these plates increase continually in importance.

Summary. The AT and BT blanks 1.0 to 1.2 mm in thickness sawed from a quartz crystal are lapped off in three stages with successively finer abrasive to a thickness of 0.3 to 0.8 mm, according to the desired resonance frequency. This operation can be performed mechanically with the planetary lap machine with which 30 to 55 plates are lapped simultaneously. At the same time the plates automatically take on the contour necessary for good "activity". With six of these machines three lots of crystals can be finished in less than 15 minutes of actual running time. The moment at which the plates have reached the desired thickness is determined by measuring the resonance frequency of the crystal while the machine is running; this is possible because of the fact that the crystals are continually brought into vibration by the lapping process, the piezoelectric effect thereby causing a weak alternating voltage on the lap plates. By a transposition of the crystals of a lap at a certain moment in the third lap stage the frequency spread of the lapped plates is kept within about 15 kc/s (at 8 Mc/sec). After lapping, at least 1 micron is etched off from the crystals in a water solution of ammonium bifluoride. Each crystal is then adjusted accurately to the nominal frequency by further individual etching. Crystals finished in this way are practically free of ageing phenomena.

---

[8]) A. C. Prichard, M. A. A. Druesne and D. G. Mc Caa, Acta Cryst. 3, 73, 1950 (No. 1).

MAKING LAP PLATES FOR THE MANUFACTURE OF QUARTZ OSCILLATOR-PLATES



Crossed serrations, in which the abrasive can accumulate, being milled into the meehanite plates of the planetary lap machine used for the mass manufacture of quartz oscillator-plates (see the article on that subject in this number).

MEASURING THE DEIONISATION TIME OF GAS-FILLED DIODES AND TRIODES

by K. W. HESS.

621.314.67:621.385.38:537.567.569

The charge of the ions compensating the space charge formed by the electrons in a gas-filled discharge tube is neutralized mainly on the walls and the electrodes of the tube. After the current has ceased to flow it takes a certain time for the tube to become free of ions again. If the tube is to be used periodically this finite deionisation time sets a limit to the permissible frequency.

Introduction

In a vacuum tube, if the anode voltage is not abnormally high, the current is determined mainly by the space charge of the electrons. In gas-filled tubes this space charge is neutralized by the positive gas ions, so that at low voltages it is possible to send heavy currents through these tubes. In this article we shall devote our attention particularly to the influence that ions have on the properties of gas-filled diodes and triodes (also called relay valves or thyratrons); we shall not deal with the phenomena occurring, for instance, in "TL" lamps, neon tubes, etc.

So long as the voltages between the electrodes are not high enough to bring about a gas discharge the passage of current through a gas-filled tube is comparable to that in a vacuum tube, but as soon as the anode voltage exceeds a certain value, the ignition voltage, the tube begins to function as a gas tube. In a diode this ignition voltage has an almost constant value, whilst in a triode it is greatly dependent upon the grid voltage. This means that in the case of a triode for every value of the anode voltage (provided it is not too low) there is a certain critical grid voltage above which the tube ignites. Contrary to the case of a vacuum triode (except for some rare cases), after the ignition has taken place the grid voltage has practically no influence upon the current passing through the tube. The reason for this is that when the grid is negative a space charge of positive ions is built up around it and neutralizes the negative potential.

After the current has ceased to flow not all the gas ions will immediately recombine. In general the rule is that the number of ions σ in the tube decreases with the time t according to an exponential function:

$$\sigma = \sigma_0 \, e^{-\frac{t}{\tau_0}}.$$

The characteristic time τ_0 in this formula may be called the deionisation time. It depends, of course, upon the possibilities present for the ions to lose their charge. This takes place mainly on the walls and electrodes of the tube, and it will therefore be greatly affected by the manner on which the tube is constructed. The less positive, or, as often occurs, the more negative the voltage at one or more of the electrodes, the shorter is the deionisation time, whilst on the other hand the greater the preceding anode current the longer is the deionisation time. Finally τ_0 increases with increasing gas pressure [1]).

What is of importance in practice is not the deionisation time as defined above but rather the effective deionisation time, τ_{eff}, which is the period of time that has to elapse, after the current has ceased to flow, until the working of the tube is no longer affected by the remaining ions. Obviously, τ_{eff} will have a different value for each different application of the tube [2]). The deionisation time will be particularly noticeable when the tube is working periodically, as for instance in a rectifying circuit or in a relay circuit, when at every ignition ions may still be present in the tube owing to the preceding flow of current through it.

Let us first consider the case of a diode. As already indicated, the tube ignites when the anode voltage reaches a certain value. In many cases this ignition voltage depends upon the number of ions in the tube, being lower the more ions there are. When for a particular application the ignition voltage is required to be above a certain value then it

---

[1]) For this subject see, e.g., H. B. de Knight, Proc. Inst. Electr. Engrs 96, Part III, 257, 1949.

[2]) In literature all sorts of definitions are given for the deionisation time, but mostly it is the effective deionisation time that is meant, in some form or other. The essence of the conception of deionisation time is certainly implied in our original definition.

is necessary to ensure that the interval of time between two successive ignitions is so large that the ignition voltage for the second ignition, lowered by the presence of ions due to the first passage of current, does not fall below that prescribed level. In other words, there is an effective deionisation time setting an upper limit to the frequency at which the tube can be worked.

A similar effect arises in the case of a triode. The space-charge sheath formed around the grid while current is passing does not immediately disappear after the current ceases to flow. Consequently the tube cannot be controlled by means of the grid potential until a certain effective deionisation time has elapsed. When a sufficiently high anode voltage is applied before that time has expired the tube will reignite while the grid voltage is still below the critical voltage corresponding to that anode voltage. Thus there is again an upper limit set for the frequency at which the ignition of the tube can be controlled with the aid of its grid voltage.

For all kinds of rectifying tubes, both diodes and triodes, the finite deionisation time in fact increases the risk of backfire in the tube: when the anode voltage is strongly negative while there are still many ions in the tube it is quite easy for a glow (sometimes arc) discharge to take place, causing electrons to travel from the anode to the cathode. For this effect, too, an effective deionisation time can be defined, which is generally much shorter than the deionisation times referred to above. However, we shall not consider this phenomenon any further here.

In this article we shall discuss some methods for measuring the deionisation time in various cases taken from practice. We could carry out the measurements with a variable frequency and determine at what frequency the tube ceases to function properly, but this is impracticable on account of the usually high power consumption and the technical difficulty of arranging a measuring set-up for any desired frequency. We prefer to use a system producing a condition which, when working at the mains frequency, is comparable to the working of the tube with the desired frequency, then determining τ with the aid of an oscillograph.

Measurement of the deionisation time in a triode by means of the grid current

First we shall describe a method for determining easily and quickly the order of magnitude of the deionisation time in a triode, using for that purpose a direct indicator of the number of positive ions in the tube, namely the grid current. An alternating voltage is applied to the anode circuit of the triode to be tested, while the grid is connected via a resistor to a negative voltage source. The negative voltage is chosen high enough to ensure that the tube cannot of itself ignite during the positive half-cycle of the alternating voltage on the anode circuit. At a moment t_1 during this half-cycle the grid receives a voltage impulse from a peak transformer such as described, for instance, in an article recently published in this journal [3]), the value of this impulse being so chosen as to cause the tube to ignite. Current then begins to flow and the anode voltage drops to the level of

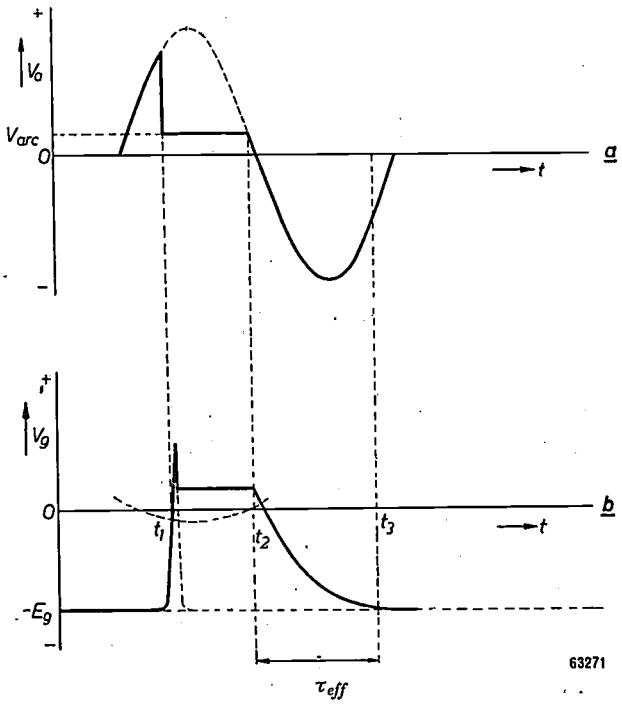


Fig. 1. The variation of: *a*) the anode voltage, *b*) the grid voltage of a gas triode fed with alternating current. At the moment t_1 the grid, which normally has a negative bias $-E_g$, receives a voltage impulse causing the tube to ignite. The anode voltage then drops to the arc voltage level V_{arc}. At the moment t_2 the tube is extinguished on account of the anode voltage dropping below V_{arc}. During the working period the grid voltage is kept at a certain level as a consequence of the positive ion current (this level being generally somewhat lower than the arc voltage). After extinction of the discharge the ions gradually disappear and the grid voltage gradually drops to the level $-E_g$, which is reached at the moment t_3. Thus the interval of time t_3—t_2 is a measure of the deionisation time.

the arc voltage V_{arc} (*fig. 1a*). When at the moment t_2 the alternating voltage drops below the arc voltage the tube extinguishes. The grid voltage, which after the very short impulse would tend to

[3]) K. W. Hess and F. H. de Jong, Controlling the luminous intensity of fluorescent lamps with the aid of relay valves, Philips Techn. Rev. **12**, 83-93, 1950 (No. 3).

drop again to the original negative value, is kept at a level usually positive and somewhat lower than the arc voltage, owing to the voltage drop in the grid resistor brought about by the ion current flowing to the grid. Thus the voltage at the grid is a measure of the ion current. After the extinction of the tube ions continue to be present in it for some time, though diminishing in number, and the grid voltage gradually drops to its original negative value (fig. 1*b*), which is reached at the moment t_3. Thus t_3—t_2 is a measure of the deionisation time.

By keeping the frequency of the A.C. anode voltage low (the mains frequency of 50 c/s is highly suitable) the moment t_3 can be made to fall within the negative half-cycle of that voltage. The whole phenomenon is repeated in the next cycle of that voltage, the tube striking at the moment t_1' when the grid receives the next voltage impulse. By displaying the grid voltage on the screen of an oscillograph it is quite easy to determine the order of magnitude of the deionisation time. Some oscillograms obtained in this way have been reproduced in *fig. 2*.

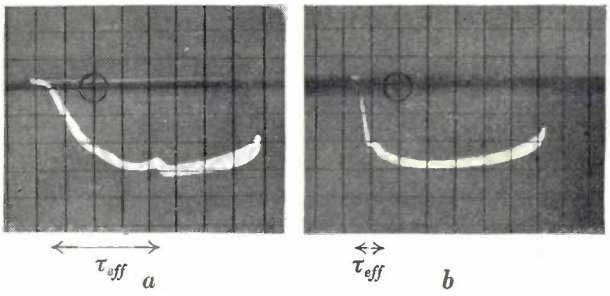


Fig. 2. Two oscillograms showing the variation of the grid voltage, *a*) for a tube with long deionisation time (about 4 msec), and *b*) for a tube with short deionisation time (about 1 msec). The part of the curve denoting the variation of the grid voltage prior to the moment t_2 (cf. fig. 1) is not included in these oscillograms.

To determine accurately the effective deionisation time for a specific case occurring in practice we must follow other methods, some of which will be described below.

Measurement of the effective deionisation time

Deionisation in a triode

We shall now deal with the measurement of the effective deionisation time which in the case of a triode determines what period of time has to elapse after the current ceases to flow before the grid is again able to prevent ignition. Here we use the set-up illustrated diagrammatically in *fig. 3*.

The heavily lined current circuit ("low-voltage circuit") serves for the production of the ions and the heating of the tube G under test to its working temperature; the transformer T_2 supplies a voltage (between 30 and 60 V) just high enough to send the desired current through the circuit formed by

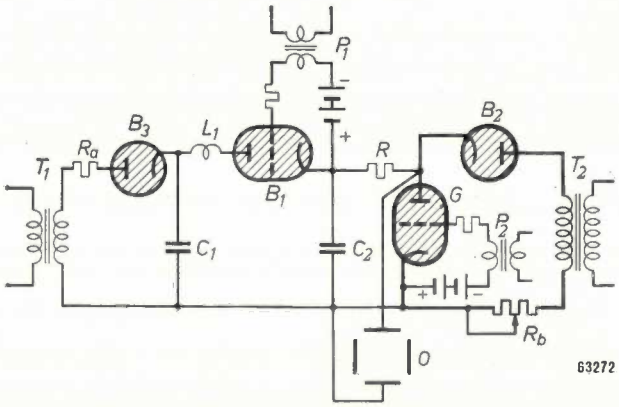


Fig. 3. System for measuring the effective deionisation time of a triode. The manner in which it functions is described in the text. The right-hand part of the diagram T_2-B_2-G-R_b supplies G, the tubeu nder test, with its correct anode current, thus maintaining it at its operating temperature. The left-hand part of the diagram shows the system for producing positive voltage pulses with steeply ascending front. The voltage across G is displayed on the screen of an oscillograph.

T_2, B_2, G and R_b. The function of the gas-filled diode B_2 will be made clear presently; R_b is a load resistor. The grid of G is connected to a negative-voltage source; connected to the grid circuit is also a peak transformer P_2. At some time during the positive half-cycle of the anode voltage of G, say at the moment t_1 (*fig. 4a*), the grid receives a voltage impulse causing G to ignite, and the anode voltage drops to the arc-voltage level (the rest of the voltage is taken up by R_b). At the moment t_2 the anode voltage drops below the arc-voltage level and G extinguishes. During the negative half-cycle the voltage appearing across G is determined by the very low conductance of G in the extinguished state and of the diode B_2 (likewise extinguished). To a first approximation we may say that this voltage is zero. The whole process is repeated during the next cycle.

Now we suppose that at the moment t_3 (during the negative half-cycle of the transformer voltage) the gas-filled triode B_1 is ignited by an impulse reaching its grid from the peak transformer P_1. The charge from the capacitor C_1, forming part of the rectifying circuit T_1-R_a-B_3-C_1 ("high voltage circuit"), is thereby caused to flow through the self-inductance L_1 and through B_1 to C_2, thereby charging C_2; with the right selection of C_1, L_1 and C_2 this charging can be made to take place very quickly. B_1 automatically extinguishes again as soon as the charging of C_2 is ended. The anode of

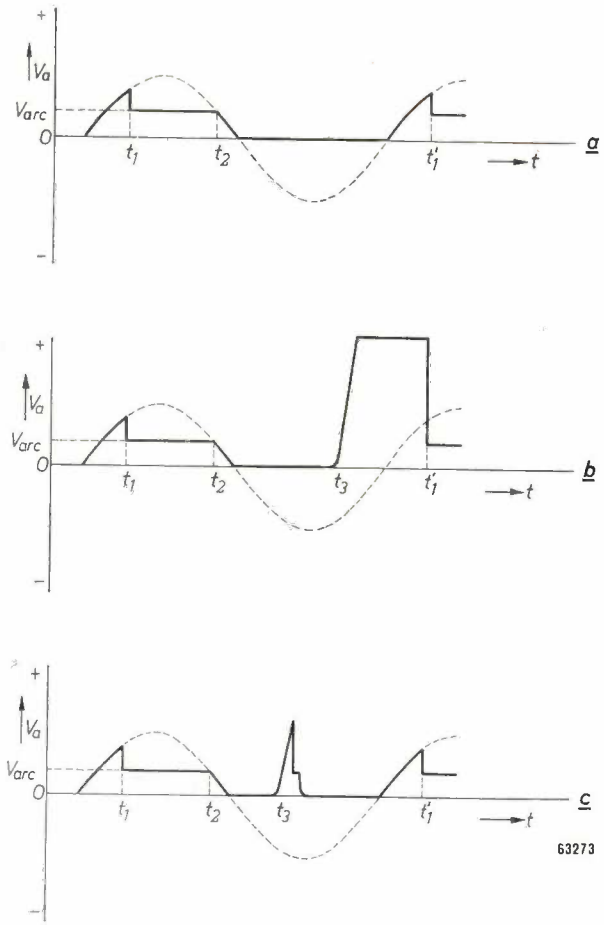


Fig. 4. *a*) Variation of the anode voltage of G in the system of fig. 3 when there is no voltage across the capacitor C_2. At the moment t_1 the tube G is ignited by an impulse on its grid, whilst at the moment t_2 it is extinguished owing to the anode voltage dropping below the arc voltage. The fact that there is no negative voltage across G during the negative half-cycle of the transformer voltage is due to the presence of the diode B_2 (see fig. 3). The process is repeated in the next cycle. *b*) As in (*a*) but during the negative half-cycle the tube B_1 is fired at the moment t_3, as a result of which a positive voltage appears across G; G does not ignite, however, because its grid voltage is negative and there are few ions in the tube, since t_3—t_2 is longer than the effective deionisation time. The voltage does not then drop again to V_{arc} until the moment t_1' when the grid of G receives another impulse and G ignites. *c*) As (*b*) but with t_3—t_2 shorter than the effective deionisation time. The tube G ignites owing to the presence of sufficient ions, in spite of the low grid voltage, as soon as the anode voltage reaches a certain value. This value depends upon the number of ions, and thus upon t_3—t_2. The voltage across G then drops within a few μsec to the arc voltage level and, after completion of the discharge of C_2, to zero.

G thus receives a voltage impulse with a steeply ascending front, as sketched in fig. 4*b*. The amplitude of the voltage supplied by T_1 (some hundreds of volts) and the values of the circuit elements C_1, L_1 and C_2 are so chosen that the height of this pulse is greater than the voltages generally occurring in practice, but not so great that, in the absence of ions in the tube, G will strike at the given negative grid bias. When the aforementioned condition is fulfilled the capacitor C_2 will therefore not be able to start discharging until G ignites owing to excitation of its grid by an impulse, i.e. at the moment t_1'.

If, however, the moment is so chosen that there are still ions in G when C_2 is being charged, then G will ignite, without a grid impulse, when the anode voltage — i.e. the voltage across C_2 — exceeds a certain value. The charge from C_2 will then flow off directly through G. The voltage that has to lie across G to cause the tube to ignite obviously depends on the number of ions in the tube; thus the shorter the interval t_3—t_2 the lower is the peak voltage that appears across C_2 (fig. 4*c*).

Since the phenomenon is repeated at the mains frequency the voltage across G can be displayed on the screen of the oscillograph and we can thus easily determine at what interval of time the full voltage across C_2 can still be borne by G. To do this we have to vary the phase shift between the impulses from P_1 and P_2. This can be done in various ways. In our experiments we used an induction regulator, consisting of an induction motor with fixed rotor and the stator connected to three-phase mains. The primary current for one of the peak transformers is taken from the rotor winding and by turning the rotor by hand it is possible to shift the phase of this current with respect to the stator phase.

The minimum interval of time between the extinction of G and the impulse on C_2 at which the tube G still does not ignite is the effective deionisation time sought. Some oscillograms of the phenomenon are represented in *fig. 5*, which clearly show that as the interval of time is reduced the height of the impulse sufficient to cause G to ignite is also reduced.

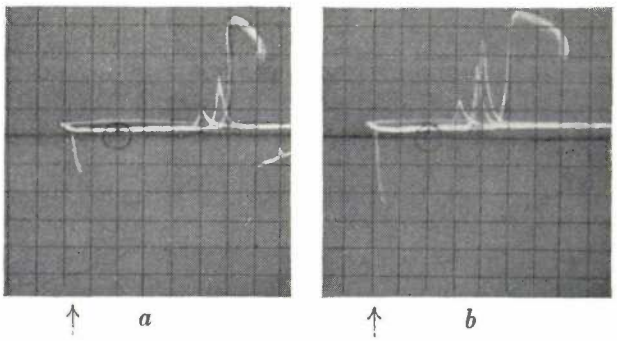


Fig. 5. Two oscillograms of the phenomenon described in fig. 4, *a*) for a tube with long τ_{eff} (1.2 msec), *b*) for a tube with short τ_{eff} (0.85 msec). Only the variation of the anode voltage in the interval of time from t_2 up to shortly after t_3 is represented, the curves for different intervals t_3—t_2 being shown in the same oscillogram. It is clearly seen how the height of the impulse at which G just ignites varies with this interval of time. The descending slope of the impulse is not discontinuous as in fig. 4*c* but gradual, owing to the influence of the measuring potentiometer. The arrow pointing to the small vertical line on the left of the oscillograms indicates the moment t_2.

The purpose of the tube B_2 in the heavily lined part of the system is to prevent the charge from C_2 flowing away through the secondary winding of T_2 when the tube G is not ignited.

Since B_2 prevents the anode of G from becoming negative (with a negative anode the ions, of course, recombine more rapidly), this set-up furnishes a very safe velue for the deionisation time.

As the definition of the effective deionisation time shows, this time varies according to the practical application of the tube. By making a small alteration to the "low-voltage circuit" of the system according to fig. 3, we can adapt this for measuring the effective deionisation time of a triode used in a polyphase rectifying circuit. In such a circuit there are several rectifying tubes (here triodes), with the current passing through each of them in turn. When commutating, the current flowing through a tube very rapidly drops from a fairly high level to zero [4]). In this case a longer effective deionisation time is to be expected than in the case of a single-phase rectifying circuit. The alteration made to the "low-voltage circuit" of fig. 3 is illustrated in *fig. 6* and explained in the subscript.

For tubes not passing too high a current the system of fig. 3 can be simplified, the working current being supplied directly by the "high-voltage circuit", as illustrated in *fig. 7*, where the tube under test is again represented by G. Upon the gas triode B being fired the capacitor C is charged by the direct-current source A. The grid bias of G is such as to prevent this tube from igniting. When the grid of G receives a voltage impulse from the peak transformer P_2 the tube will ignite and C is discharged through G. This process can be made to take place very quickly by choosing suitable values for L and C. It is then possible to re-ignite B, and C is recharged by applying an impulse to the grid of B.

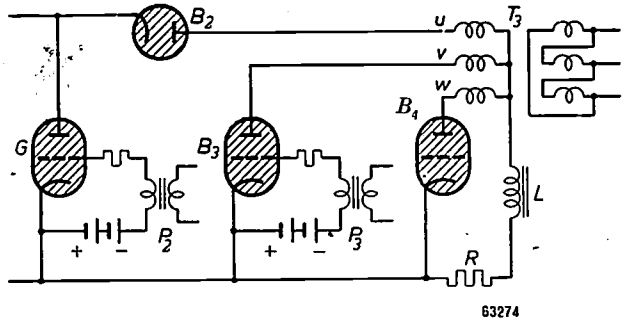


Fig. 6. The "low-voltage part" of the system for the measurement of the effective deionisation time in a triode G contained in a polyphase (here three-phase) rectifying circuit. The tubes G, B_3 and B_4 are connected to the successive phases u, v and w of the three-phase transformer T_3. In the part of the cycle when the anode of G is positive G is ignited by a voltage impulse from the peak transformer P_2 applied to the grid (which is normally so heavily biassed as to prevent G from striking of itself). It is known that in such a circuit two of the tubes cannot conduct simultaneously (see the article quoted in footnote [4])). B_3 is now ignited by an impulse applied to its grid from the peak transformer P_3 at a moment when the voltage of phase u has not yet dropped below the arc voltage of G but the voltage of phase v already exceeds the voltage of phase u. As soon as B_3 is ignited the current through G very quickly drops from a fairly high value to zero, and this is just the condition under which it is desired to measure the effective deionisation time. By varying the moment of ignition of B_3 it is possible to change the conditions under which the effective deionisation time of G is measured.
The tube B_4 (the grid of which is so heavily biassed that the tube normally strikes as soon as the voltage of phase w exceeds a certain value) serves only for ensuring a practically constant value of the current supplied by T_3.
B_2 again prevents the capacitor C_2 (cf fig. 3) from discharging through the transformer.

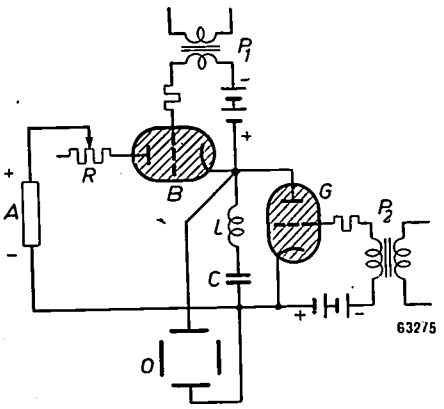


Fig. 7. Simplified system with which the same object can be achieved as with that of fig. 3. Here there is no separate circuit for supplying the operating current for the tube.

If the interval of time between the discharge of C through G and its recharging is longer than the effective deionisation time then C will remain charged, but if that interval of time is shorter then C will discharge through G as soon as the voltage reaches a certain value. Thus it is possible to determine the effective deionisation time by measuring with the aid of an oscillograph the voltage across G as a function of the interval of time (variable with a phase regulator) between the impulses on the grids of G and B.

In this case the anode of G is negative during part of the cycle owing to the influence of L.

Deionisation in the case of a diode

We shall now describe an oscillographic method of measuring the effective deionisation time determined by the decrease of the ignition voltage of a diode, taking as example a diode used in a relay circuit. For the proper functioning of such a circuit it is necessary for the ignition voltage of the diode to exceed a certain value. The case chosen is that of a glow-discharge tube with cold cathode, but the method to be described applies equally well for a diode with heated cathode. The circuit is given

[4]) Tj. Douma, Voltage impulses in rectifiers, Philips Techn. Rev. 9, 135-146, 1947.

in *fig. 8*, where the heavily lined circuit represents an imitation of the relay circuit. The principle on which this works is as follows. By means of two identical networks, each containing a controlled

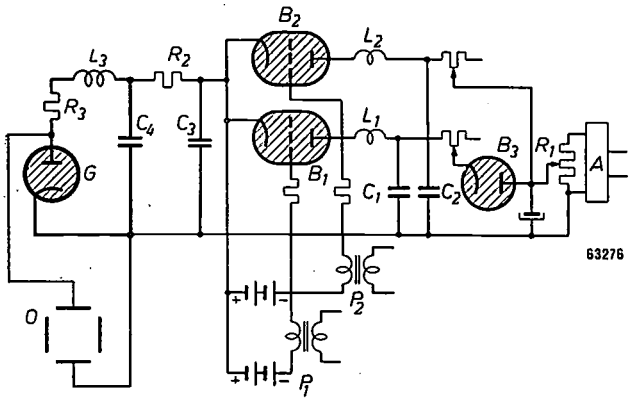


Fig. 8. System for measuring the deionisation time of a diode G in a relay circuit. The choke L_3 and the resistor R_3 form an imitation of a relay. C_3, R_2 and C_4 determine the form of the impulse, which has to resemble as closely as possible the shape of the impulse under normal working conditions. The system functions in the following way. When a positive voltage impulse (from the peak transformer P_1) is applied to the grid of the gas triode B_1 this tube ignites and the charge of the capacitor C_1 is transmitted to C_3 and thence to C_4. When the voltage across C_4 is high enough G ignites, C_4 and C_3 then being discharged, and G is extinguished again. After some time B_2 is ignited by an impulse from the peak transformer P_2. B_2 and B_1 work in exactly the same way (the circuits $B_1L_1C_1C_3$ and $B_2L_2C_2C_3$ are identical) and thus the process is repeated, G being ignited and extinguished again. If as a result of the first ignition there are ions in the tube G at the moment of the second ignition then the ignition voltage will be lower the second time and the peak voltage of the impulse across C_4 will not have the maximum value. By measuring this peak voltage as a function of the interval of time between the two impulses the deionisation time of the tube G is determined. The easiest way to do this is to make the phase of one of the peak transformers (P_2 in the diagram) variable with respect to that of P_1 by means of a phase regulator (see the text) and then to apply the anode voltage of G to the vertical deflection plates of an oscillograph.
The tube B_3 serves to prevent the capacitor C_1 being charged by C_2 when the interval between the impulses is short. The whole system works with a frequency of 50 c/s.

gas triode, positive voltage pulses can be applied to the anode of the tube G under test, with variable interval of time between the two pulses. The first pulse causes the tube to ignite, and after the discharge of the capacitors C_4 and C_3 the voltage across the tube drops below the arc voltage and the tube is extinguished. When the anode receives the next pulse the tube is reignited and the process is repeated. If ions were still present in the tube at the moment of the second pulse then the ignition voltage would be lower than that for the first pulse and the measured peak voltage of the second pulse would be lower than the normal ignition voltage (*fig. 9*). By varying the interval of time between the two pulses we can follow the variation of the peak voltage for the second pulse and thus determine the time which has to elapse between the two pulses for the peak voltage of the second one to be just as high as that of the first. This is the deionisation time. The whole phenomenon can be seen on the screen of the oscillograph. A description of the system employed is given in the subscript to fig. 8.

It must be pointed out that in order to get an exact measurement of the deionisation time it would really be necessary to determine the time elapsing between the moment that the current due to the first pulse ceases to flow through G and the moment that the voltage begins to rise as a result of the second pulse, ignoring the "rise time" of that voltage. Now from *fig. 10*, sketching the variation of the current passing through G during an impulse, it is obvious that it would be very difficult to determine exactly at what moment

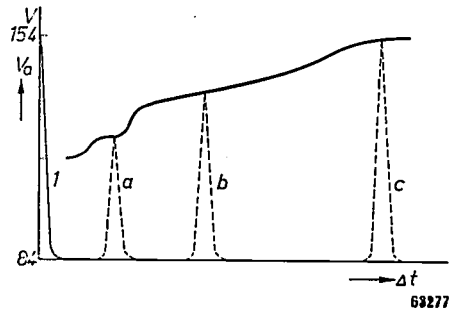


Fig. 9. Variation of the peak voltage of the second impulse at the anode of G (in the system according to fig. 8) as a function of the interval of time Δt between two impulses. When there are no ions present the ignition voltage of the tube is 154 V and the operating voltage is 84 V. Curve *1* is the variation of the anode voltage of G due to the first impulse. The dotted impulses *a*, *b* and *c* indicate how the anode voltage varies during the second impulse. The maximum value of this voltage depends upon the interval of time between the second and the preceding impulse. The heavily drawn curve shows how the ignition voltage at the second impulse varies according to that interval of time.

the current actually ceases to flow, and that is why we prefer to measure the interval of time between the beginning of the first impulse and the beginning

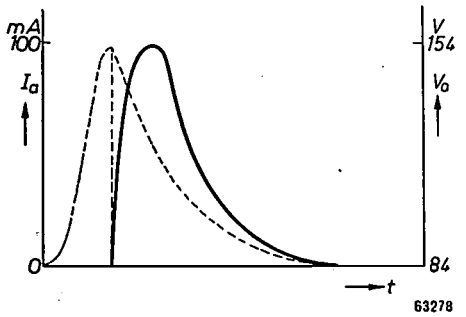


Fig. 10. Variation of the current passing through a glow-discharge tube with cold cathode during a voltage impulse on the anode, plotted as a function of time. The moment at which the current ceases to flow cannot be exactly determined. The variation of the anode voltage is indicated by a dotted line.

of the second one. For all practical purposes this is the most important interval of time; in this way the maximum switching frequency at which the tube can function is determined directly.

Summary. Effective deionisation time is defined as the time required, after current has ceased to flow, for the ions in a gas-filled tube to be so far recombined as no longer to have any adverse effect upon the working of the tube. It is useful to be able to measure this effective deionisation time in cases where a gas-filled tube has to be switched on and off several times per second. The deionisation time determines the maximum frequency for which the tube is suitable. Descriptions are given of: a method for determining quickly the order of magnitude of the deionisation time in a triode, a system for determining the effective deionisation time of a triode used, inter alia, in a polyphase rectifying installation, and a method for determining the effective deionisation time of a diode in a relay circuit.

BOOK REVIEW

Application of the Electronic Valve in Radio Receivers and Amplifiers. Volume 1: R.F. and I.F. amplification - Frequency changing - Determining the tracking curve - Parasitic effects and distortion due to the curvature of valve characteristics - Detection, by B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen; 416 pages, 256 illustrations. — Published by N.V. Philips' Gloeilampenfabrieken, Technical and Scientific Literature Department, Eindhoven, Netherlands, 1950.

This book belongs to the series of books on electronic valves published in Philips Technical Library. Two other volumes are in preparation. Volume 2 will deal with A.F. amplification, power amplification, inverse feedback and power supply; volume 3 with control devices, stability and instability of circuits, parasitical feedback, interference phenomena and calculations of receivers and amplifiers. In this book only amplitude-modulated signals are considered. The reader is presumed to be acquainted with the theory of the electronic valve itself and to have some general knowledge on radio receivers. In the more than 400 pages the subjects covered by the title are dealt with in detail, and full calculations of many circuits are given. A list of symbols at the beginning of the book is helpful in understanding an arbitrary section of the book that might be chosen for study. An extensive contents helps the reader to find the topics in which he might be particularly interested. At the end of each part a bibliography is included referring the reader to some of the more important articles on the subjects covered by the part.

The part on R.F. and I.F. amplification starts with a discussion of the single tuned circuit, followed by an extensive treatment of two coupled circuits. Special attention is paid to the various ways of coupling an aerial to a tuned circuit and of attaining a variable bandwidth in the I.F. amplifier. Circuits for image suppression are not included. In the part on frequency changing the properties of oscillator circuits are dealt with extensively. Separate sections are devoted to squegging oscillation and to frequency drift. This is followed by a separate part on the determination of the tracking curve. The next part deals with effects due to the curvature of valve characteristics: cross-modulation, modulation distortion, modulation hum, whistles. In the part on detection most attention is paid to diode detection, which is fully treated. The result of a difference between D.C. and A.C. resistance in the diode circuit and the reaction of the diode circuit on the preceding tuned circuit are clearly elucidated. All parts are written clearly and the reader is given full information on any of the subjects treated in the book.

B. D. H. Tellegen

of the second one. For all practical purposes this is the most important interval of time; in this way the maximum switching frequency at which the tube can function is determined directly.

Summary. Effective deionisation time is defined as the time required, after current has ceased to flow, for the ions in a gas-filled tube to be so far recombined as no longer to have any adverse effect upon the working of the tube. It is useful to be able to measure this effective deionisation time in cases where a gas-filled tube has to be switched on and off several times per second. The deionisation time determines the maximum frequency for which the tube is suitable. Descriptions are given of: a method for determining quickly the order of magnitude of the deionisation time in a triode, a system for determining the effective deionisation time of a triode used, inter alia, in a polyphase rectifying installation, and a method for determining the effective deionisation time of a diode in a relay circuit.

BOOK REVIEW

Application of the Electronic Valve in Radio Receivers and Amplifiers. Volume 1: R.F. and I.F. amplification - Frequency changing - Determining the tracking curve - Parasitic effects and distortion due to the curvature of valve characteristics - Detection, by B. G. Dammers, J. Haantjes, J. Otte and H. van Suchtelen; 416 pages, 256 illustrations. — Published by N.V. Philips' Gloeilampenfabrieken, Technical and Scientific Literature Department, Eindhoven, Netherlands, 1950.

This book belongs to the series of books on electronic valves published in Philips Technical Library. Two other volumes are in preparation. Volume 2 will deal with A.F. amplification, power amplification, inverse feedback and power supply; volume 3 with control devices, stability and instability of circuits, parasitical feedback, interference phenomena and calculations of receivers and amplifiers. In this book only amplitude-modulated signals are considered. The reader is presumed to be acquainted with the theory of the electronic valve itself and to have some general knowledge on radio receivers. In the more than 400 pages the subjects covered by the title are dealt with in detail, and full calculations of many circuits are given. A list of symbols at the beginning of the book is helpful in understanding an arbitrary section of the book that might be chosen for study. An extensive contents helps the reader to find the topics in which he might be particularly interested. At the end of each part a bibliography is included referring the reader to some of the more important articles on the subjects covered by the part.

The part on R.F. and I.F. amplification starts with a discussion of the single tuned circuit, followed by an extensive treatment of two coupled circuits. Special attention is paid to the various ways of coupling an aerial to a tuned circuit and of attaining a variable bandwidth in the I.F. amplifier. Circuits for image suppression are not included. In the part on frequency changing the properties of oscillator circuits are dealt with extensively. Separate sections are devoted to squegging oscillation and to frequency drift. This is followed by a separate part on the determination of the tracking curve. The next part deals with effects due to the curvature of valve characteristics: cross-modulation, modulation distortion, modulation hum, whistles. In the part on detection most attention is paid to diode detection, which is fully treated. The result of a difference between D.C. and A.C. resistance in the diode circuit and the reaction of the diode circuit on the preceding tuned circuit are clearly elucidated. All parts are written clearly and the reader is given full information on any of the subjects treated in the book.

B. D. H. Tellegen

VOL. 12 No. 7, pp. 185-212 JANUARY 1951

Philips Technical Review

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

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

THE "PHOTOFLUX" SERIES OF FLASHBULBS

by G. D. RIECK and L. H. VERBEEK.

771.448.4

It is known that right in the beginning of photography, about 100 years ago, photographs were already being taken with a flash of light. This is probably even the oldest form of artificial-light photography, since with the means of permanent artificial light available in those days and the very poor sensitivity of the sensitive plates of that time, the exposures would certainly have been unduly long. The flash was obtained by burning i.a. magnesium powder.

Nowadays flashlight technique has reached a very high degree of refinement, as will be shown in this article.

Introduction

It was about 20 years ago that the "Photoflux" and similar flashbulbs were introduced to replace the old flashlight produced by combustion of a mixture of magnesium powder and an oxidizer in air. Their success was remarkable. Only in one application, for "magnesium flash bombs " when photographing from aircraft, is the "open" flashlight still in use, but otherwise it was very soon ousted by the "closed" flashlight. "Ousted" is not really the right word, since with their safe, clean and perfectly controlled working the modern flashbulbs have not only replaced the open flashlight but by their introduction have brought about a vast expansion in the field of flash photography. The flashbulb has become the inseparable companion of the press photographer — probably the foremost user of this article — and also in amateur photography it has taken a very important place. Many are the accessories that have in consequence been brought on the market for flash photography, such as reflectors and synchronizers, and many are the types of cameras that have been fitted with built-in synchronizers with contacts for one or more flashbulbs. In this connection mention may be made of the Philips flash camera, which has not only a synchronizer but also a lamp socket and reflector built into it [1]).

The principle of the "Photoflux" lamp was dealt with in this journal as far back as 1936 [2]). The principle has not been altered much, nor is there much difference in the appearance of the present-day lamp compared with that described at the time. The properties of the lamp however have been improved or changed in a great many respects, so that it is deemed opportune to discuss these "Photoflux" lamps anew; they are now being produced in a series of five types, some of them in different colours. We shall not particularly dwell upon the points of resemblance and difference compared with the older types, since this is not likely to be of much interest to the reader.

General construction of "Photoflux" flashbulbs

The flash from the "Photoflux" flashbulbs is brought about by the combustion of fine wire of an aluminium-magnesium alloy (with 5-7% Mg). A certain quantity of this wire — it is only 32 μ thick — is placed loosely in a bulb like that of an incandescent lamp, and during this process it is given a sort of "permanent wave", so that it is uniformly distributed all over the inside of the bulb in the form of a loosely knit, resilient ball. The sealed bulb is further filled with oxygen under somewhat less than atmospheric pressure and contains also an ignition filament which can be heated by an electric

[1]) See Philips Techn. Rev. 12, 149, 1950 (No. 5).

[2]) J. A. M. van Liempt and J. A. de Vriend, The "Photoflux", a light-source for flashlight photography, Philips Techn. Rev. 1, 289-294, 1936.

current from an external source (see *fig. 1*). When the filament reaches a certain temperature an explosive paste applied to it ignites and the glowing particles scattered from it in turn ignite the ball of aluminium-magnesium wire at numerous points simultaneously, so that the whole ball burns away in a very short time.

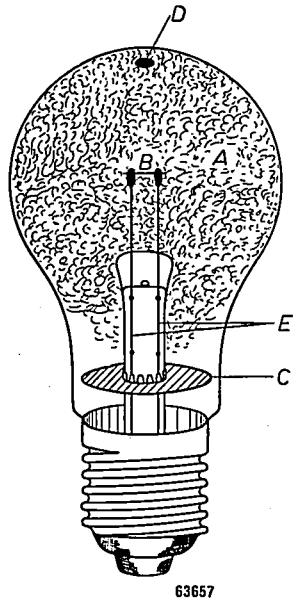


Fig. 1. Construction of a "Photoflux" flashbulb. *A* a ball of curled aluminium-magnesium wire, *B* ignition filament with explosive paste, *C* aluminium plate, *D* spot of cobal salt, *E* fuse wires.

The electric current for the ignition is usually derived from a battery, but the large types of "Photoflux" lamps can also be ignited from the lighting mains, for which purpose these types are provided with a quick-acting fuse fitted in the lamp base and forming the connection to the mains lead; owing to the arcing in the gas following upon the ignition of the lamp a very heavy current is drawn from the mains, but then this safety fuse burns through so quickly as to save the fuse of the house mains.

The burning of the Al-Mg wire causes the oxygen in the lamp to be heated quickly to a high temperature, so that for a moment — until the gas has cooled down again — the bulb comes under a pressure that may amount to several atmospheres. A normal cold bulb can withstand this pressure, but in the flashbulb the wall of the bulb is also subjected to a bombardment of glowing particles of ash which are apt to crack the glass where they strike against it, and in that weakened condition the bulb cannot withstand the pressure. The bulbs of the "Photoflux" lamps are therefore lined with a layer of lacquer in which the glowing particles are cooled down before they are able to damage the glass. Near the base of the lamp, where the layer of lacquer cannot be applied owing to the sealing of the bulb, the glass is protected against the glowing particles by an aluminium plate mounted in the neck of the lamp (see fig. 1).

External damage, such as scratches, which may easily result when, for instance, a couple of unpacked flashbulbs are carried in the coat pocket and rub against each other, may likewise so weaken the bulbs that they are unable to withstand the pressure set up inside them when they are ignited. A coating of lacquer is therefore also applied to the outside of the bulb to prevent scratching of the glass. At the same time this outer coating serves to keep the fragments together and thus render the explosion harmless in the event of the bulb bursting in spite of these precautions. For that reason the coating is made fairly thick and with a tough kind of lacquer, while it is also coloured light yellow so that it can easily be checked; this colouring has hardly any effect upon the radiation of the light.

The effectiveness of these two layers of lacquer in preventing the bulbs from bursting is evident from the random tests taken as a regular check in the manufacture of these lamps. Of the tens of thousands of lamps ignited for these tests in the course of the last few years less than one in a thousand have exploded.

This applies to lamps where the bulbs were quite intact prior to ignition. Bulbs already showing slight defects, from whatever cause, are more likely to explode upon ignition, especially so when there are small cracks in the glass through which air can penetrate. To avoid this increased risk with such lamps, a spot of cobalt salt is applied on the inside at the top of the bulb of all "Photoflux" lamps; in pure oxygen this spot is blue, but when there is the slightest trace of water vapour (and this is always present in the air) it changes to pink, thus immediately betraying the presence of air in the bulb, in which case it is not advisable to use the lamp [3]).

"Photoflux" lamps that have served their purpose sometimes show a more or less extensive blackening of the laquer lining, while in some cases a bulge is to be seen in the outer coating of lacquer. These two phenomena might give rise to some doubt as to whether the layers of lacquer are after all the right solution of the explosion problem, since, if the

[3]) In addition to the slightly greater risk of bursting, when the lamps have sucked in air there is also the risk of faulty synchronization, owing to their light-time characteristic undergoing a change; cf. the final section of this article.

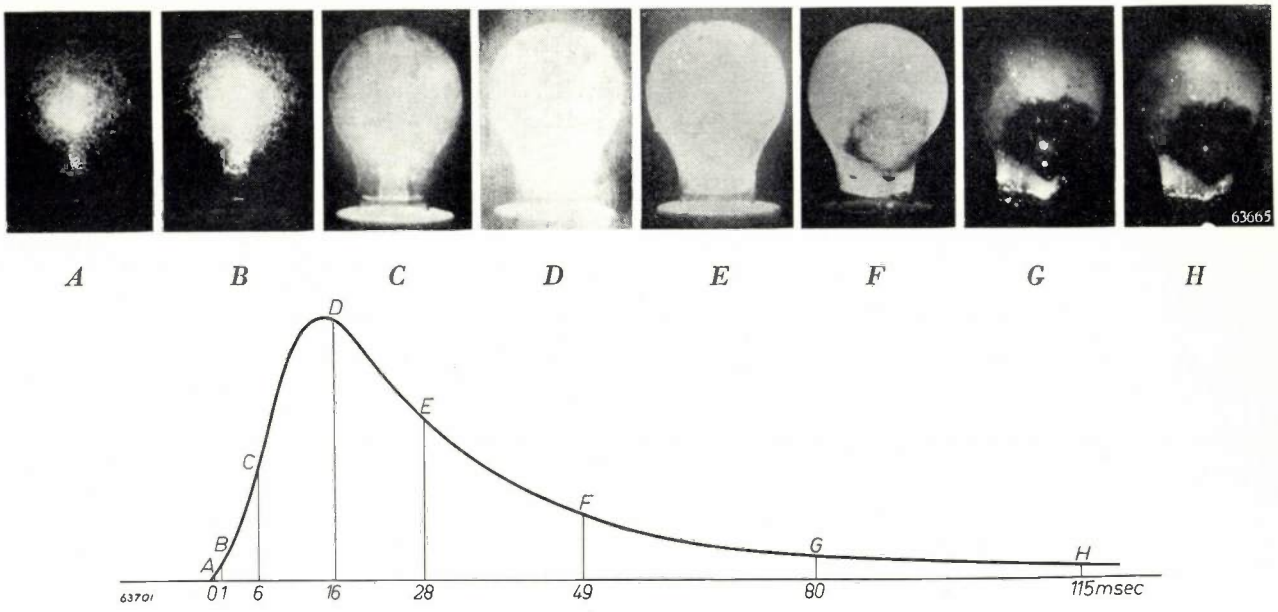


Fig. 2. Film of a burning "Photoflux" flashbulb taken at a speed of 3000 pictures per second. A number of cuts have been taken from the film corresponding to the moments indicated in the drawing (light-time characteristic) after the first picture reproduced (*A*). The film shows the formation of a black patch in the inner coating of lacquer.

blackening takes place at the beginning of the generation of the light, this will absorb some of the light, whilst as regards the bulge in the outer layer, if this is formed at the moment of maximum pressure from within, it will make the hope of sufficient support for a bursting bulb illusory.

Photographs taken with a high-speed camera (3000 pictures per second) of a series of "Photoflux" lamps in the course of being ignited have shown that there are no grounds for doubting the adequacy of the lacquer coatings. Both the blackening of the inner layer and the bulging of the outer coating take place after the main combustion and after the transmission of the light (*figs 2* and *3*). As a matter of fact, measurements of light output do not show any perceptible difference either between lamps with and lamps without an inner coating of lacquer.

The various types of "Photoflux" lamps

Fig. 4 shows the five types of "Photoflux" lamps that are now being made, the PF 110, PF 56, PF 45, PF 25 and PF 14. The numbers of this series roughly indicate the light output of the lamp in thousands of lumen-seconds (so-called photographic

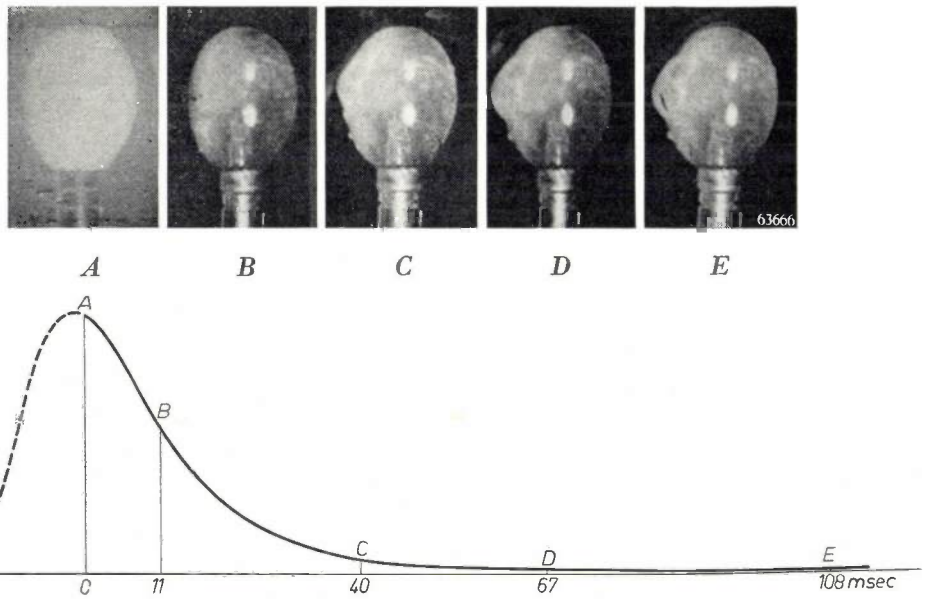


Fig. 3. Pictures taken from a film similar to that in fig. 2. This film shows the formation of a bulge in the outer coating of lacquer.

lumen-seconds, as explained below); thus the PF 56 yields 56,000 lumen-seconds [4]). To illustrate how enormous this quantity of light is, it might be mentioned that with the PF 56 lamp in a normal reflector at a distance of 4 m from the object, using a normal film and an exposure of 1/100th second, the diaphragm of the camera has to be closed down to $f/22$ to avoid over-exposure of the film.

not so very much less effective as one might suppose from the difference in the number of lumen-seconds: since the lower-powered lamps are also smaller in dimensions, with a reflector of a certain size their light is more concentrated.

The PF 25 and the PF 14 are so small that a number of them can easily be carried in a pocket. That is why the PF 25, which in spite of its small

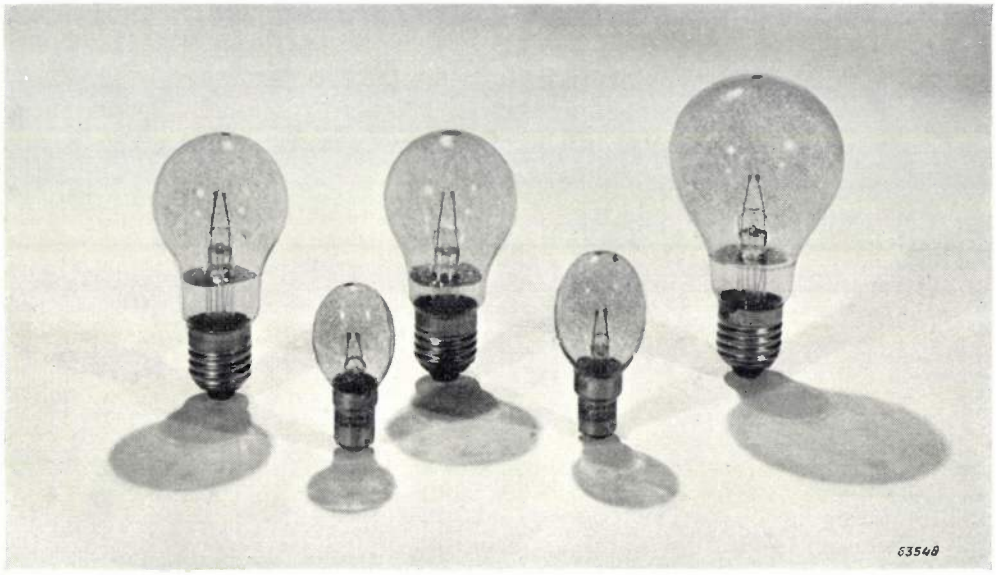


Fig. 4. The five types of "Photoflux" lamps. The PF 110 is used only in very special cases, the normal types for the professional photographer being the PF 56 and the PF 25, while the PF 14 is mainly for the amateur photographer. The PF 45 has a longer flash than that of the other types (see the last section of this article).

Such a small aperture is often undesirable, because owing to the great focusing depth also the background is brought out unpleasantly sharp. When photographing in a normal room, where it is not usually possible to get greater distances than 4 m, the PF 56 — and, a fortiori, the PF 110 — will therefore only be used when, for the sake of more artistic lighting (using a diffusing screen or illumination via the ceiling, and naturally losing a great deal of light), an abundance of light is needed. For the more usual direct illumination a smaller type mostly suffices, namely the PF 25 or the PF 14. For direct lighting these small lamps are in fact

dimensions still yields a good amount of light, has become particularly popular with press photographers. These small lamps are not fitted with a safety fuse and therefore have to be ignited with a 3-12 volt battery; they have a small bayonet type base and thus will not fit in sockets for normal lighting lamps, so that it is not likely that anyone, being unaware of this restriction, would ever try to use the lighting mains for igniting these lamps.

The type PF 45 is comparable to the PF 56 (in appearance they are even identical), but it gives a longer flash. What this exactly means and what the object of it is will be explained in the next section.

The light emitted by the types of "Photoflux" lamps mentioned here has a colour temperature of 4050 °K, which means that its spectral distribution resembles in the main that of the radiation from a black body having a temperature of 4050 °K.

[4]) It is the intention to replace the PF 56 by a flashbulb with a light output some 40 % higher than that of the PF 56 and to add to the series a new flashbulb with about 25 % less light output than the PF 56. The PF 45 will then also be replaced by a flashbulb with about 40 % higher light output. The new type number will then no longer correspond to the light output of the flashbulb.

However, the spectrum of the "Photoflux" lamps has a number of extra bands in the green; see *fig. 5* [5]). For normal black-and-white photographs a good colour rendering is obtained with this light. For colour photography, however, the spectral distribution of the light has to be different from this: daylight colour films have a spectral sensitivity that gives the right rendering of all colours only when "average daylight" is used, which means to

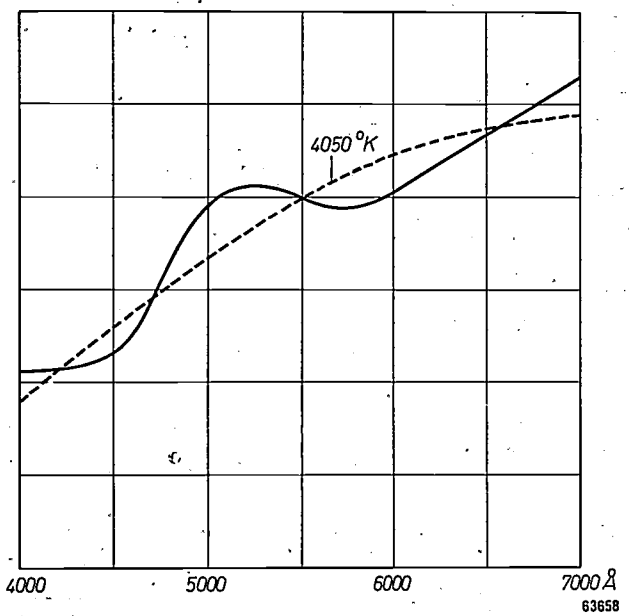


Fig. 5. Spectral distribution of the light emitted by a "Photoflux" lamp. The dotted curve represents the spectral distribution of the radiation from a black body of 4050 °K.

say that the light source has to have a colour temperature of about 5500 °K; the less commonly used artificial-light colour films are made for a colour temperature of 3200 °K or 3400 °K. For taking flash photographs on colour films some types of "Photoflux" lamps are therefore made also in two modified forms, one with a coating of blue lacquer to raise the colour temperature to about 5500 °K, the other with a yellow coating to give a colour temperature of about 3300 °K. For the blue lamps it was found necessary to choose a coating of lacquer such as would ensure sufficient absorption of the green to render the green emission bands harmless. The spectral transmission curve of the blue coating is represented in *fig. 6*, together with the spectral distribution curves of the blue lamp and of daylight, showing to what degree the colour temperatures have been matched. Of course, in the colour-temperature correction a certain loss of light has to be accepted, this amounting to about 50% with the blue lamps and about 10% with the yellow ones.

[5]) F. G. Brockman, J. Opt. Soc. Amer. 37, 652-659, 1947.

Some explanation has to be given of the figures quoted above for the total emission of light from the "Photoflux" flashbulbs as expressed in their type numbers.

When a luminous flux (or, perhaps better, a radiation flux) is measured in lumens this means that the radiation of each wavelength is given a value according to the sensitivity of the eye for that particular wavelength. For our purpose this is obviously not the right way. To give an example: the ultra-violet part of the radiation spectrum of any source of light (a rather large proportion of which passes through photographic objectives) is not included at all in the number of lumens determined in this way, because the eye is not sensitive to the ultra-violet, and yet the ultra-violet causes a considerable blackening of normal photographic material.

Furthermore, the measuring of the total radiation of the flashbulb — already in itself no simple matter — becomes rather precarious when we try to make the spectral sensitivity of the measuring instrument equal to that of the eye with sufficient accuracy. Such is evident, for instance, from the fact that when some years back the standard apparatus employed in the U.S.A. for the measurements was corrected all the numbers of lumen-seconds quoted for American flashbulbs had to be reduced by no less than 20%.

For these reasons, contrary to American usage, it is more logical give the amounts of light emitted by the "Photoflux" lamps not in visual but in photographic lumen-seconds. This unit is directly related to the desired effect of the lamp, thus not to the brightness as perceived by the eye but to the blackening of a photographic emulsion. The definition reads: A source of light yields one photographic lumen-second when, under the conditions of the D.I.N. standard specifications, it produces on an average orthochromatic emulsion the same blackening as that produced by one visual lumen-second from a sensitometer standard lamp. Measuring is now much less difficult, since instead of being obliged to imitate a very specific sensitivity curve (that of the average eye) it has been possible to fix as a standard an easily produced sensitivity curve closely approximating the said average emulsion.

The photographic lumen-second according to this definition (and the photographic lumen that is to be defined in like

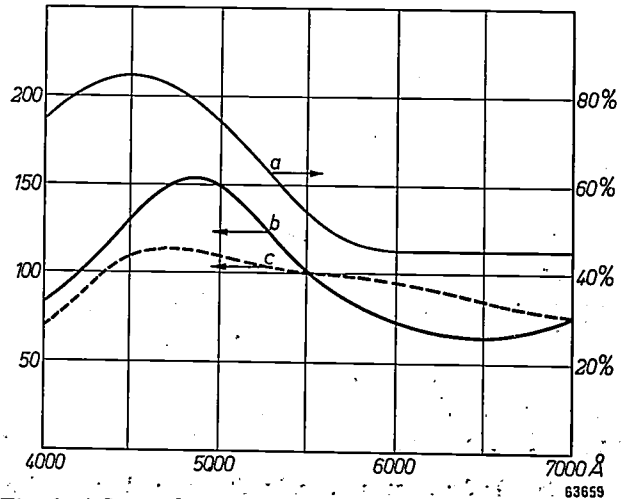


Fig. 6. *a*) Spectral transmission curve of the lacquer used for the blue "Photoflux" lamps, for colour photographs on daylight colour films.
b) Spectral distribution of the light from a blue-lacquered "Photoflux" lamp.
c) Spectral distribution of daylight (colour temperature about 5500 °K).

manner) is, it is true, related to the now more or less obsolete orthochromatic emulsion, but the numbers found prove to apply also for panchromatic emulsions to a fairly good approximation. For the "Photoflux" lamps, it is to be added, a photographic lumen-second is approximately equal to 0.7 visual lumen-second [6]).

The light-time characteristic of "Photoflux" lamps

Fig. 7 gives the average light-time curves of all "Photoflux" lamps excepting the PF 45. These curves have been produced by plotting the luminous flux (in photographic lumens) as a function of the

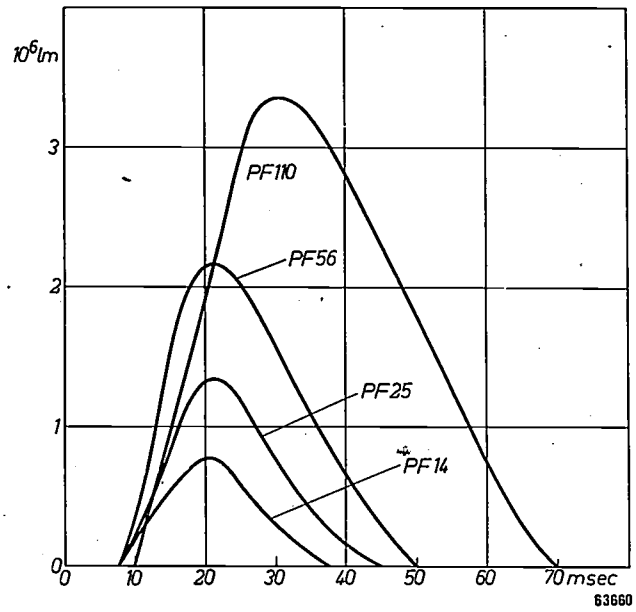


Fig. 7. Average light-time characteristics of the "Photoflux" lamps PF 110, PF 56, PF 25 and PF 14. The luminous flux (in photographic lumens) has been plotted as a function of the time (in msec) elapsing after switching on the current for the ignition filament.

time (in milliseconds) elapsing after closing the contact through which the electric current flows to the ignition filament; the manner in which the recordings are taken for these curves is described elsewhere [7]).

The area enclosed by each curve gives the total amount of radiation mentioned above. This quantity is only applicable for the effect of the flashbulb when applying the old "open-and-flash" method, where the shutter is set to *B* (ball) or *T* (time), the lens opened, the flashbulb ignited and then the lens shut again. With this primitive "synchronization method" the exposure time is not determined by the rather long time that the shutter is opened but by the duration of the flash. In the case of the PF 56, for instance, the exposure time is then about 1/40th second (disregarding the rather ineffective trailing edge of the curve in fig. 7). What constitutes an important advantage of the modern flashbulbs, however, is that they allow of a refined synchronization method whereby it is possible to photograph with much shorter exposures than the duration of the flash. Now what are the requirements that the characteristic of the lamp have to satisfy for this purpose?

In answering this question we have to distinguish between synchronization with between-the-lens type of shutters and that with focal-plane shutters.

Synchronization of flashbulb and between-the-lens shutter

In *fig. 8* curve *a* represents the (idealized) exposure characteristic of a between-the-lens shutter, where the part of the surface of the lens exposed by the shutter blades is plotted as a function of time, for instance with a set exposure of 1/100 sec. To take a photograph with this exposure time at the light from a flashbulb it will be desired that the peak of the emission (curve *b*) coincides approximately with the middle of the time interval during which the shutter is wide open. This is achieved by arranging for the shutter to be operated via the synchronizer, which first closes an electrical contact for the ignition of the flashbulb and then opens the shutter after a certain interval of time (T_0 in fig. 8).

In this case it is not the total number of lumen-seconds of the flash that is decisive for the blackening of the film but, approximately, the maximum

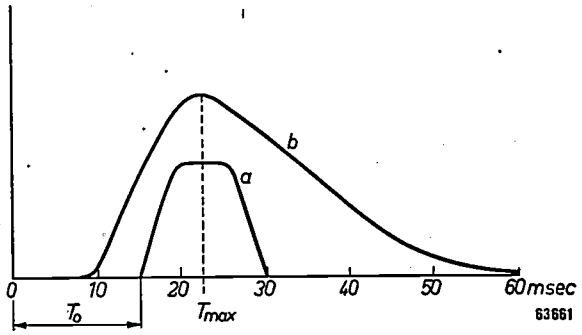


Fig. 8. *a*) Exposure characteristic of a between-the-lens shutter adjusted for an exposure of 1/100 sec. The part of the lens surface exposed by the shutter blades has been plotted as a function of time.
b) Light-time characteristic of the PF 56. The "maximum time" T_{max} is about 22 msec, like that of the PF 25 and the PF 14. The delay time T_0, i.e. the time which elapses between the closing of the contact for igniting the flashbulb and the releasing of the shutter by the synchronizer, has to be matched to the value of T_{max}.

[6]) It is only for reasons of uniformity in the denomination that the visual lumen-second is used as a unit. This argument is nevertheless likely to turn the scale, so that in the future the ratings of "Photoflux" flashbulbs will also be expressed in this unit..

[7]) See, e.g., J. A. M. van Liempt and J. A. de Vriend, Physica, The Hague 4, 354, 1937; T. H. Projector and L. E. Barbrow, Rev. Sci. Instr. **16**, 51, 1945.

luminous flux (the number of lumens in the peak of the characteristic — with the PF 56 for instance more than 2 million).

The delay time T_0 of the synchronizer has to be matched, as fig. 8 shows, to the time taken for the flashbulb to develop its maximum amount of light after the ignition contact is closed; this is called the maximum time [8]). This time should therefore be equal, within a very narrow tolerance, to one fixed value for all flashbulbs of one type, and preferably it should also be equal for different types of flashbulbs of different sizes and even of different makes, so that any of them can be used without it being necessary to change the (adjustable) delay time of the synchronizer.

At first the "Photoflux" lamps were designed for a maximum time of about 30 msec, but after the second world war and when communication was again possible with the outside world it appeared that in the meantime American manufacturers had agreed upon a shorter maximum time, viz. 20-22 msec. In fact several cameras had been placed on the market with built-in synchronizers that could only be adjusted for a maximum time of at most 25 msec.

For the sake of uniformity, and so that the "Photoflux" lamps could also be used with those cameras, the maximum time of the PF 56, PF 25 and PF 14 lamps was then likewise reduced to approximately 22 msec. This change, which in the main amounted to a quicker ignition of the aluminium-magnesium wire, was effected by changing the composition of the explosive paste on the ignition filament. As a consequence of the more rapid ignition, however, also the force of the pressure to which the bulb is subjected was increased and thus there was a greater risk of the bulb bursting. This evil has been remedied, as the figures already quoted show, by reinforcing the outer coating of lacquer.

It is perhaps well to point out that there is only question of a "refined" synchronization as referred to here when it is desired (and possible) to cut off part of a relatively too long flash with a mechanical shutter. With the gas-discharge flashtubes which have begun to come to the fore in recent years [9]) the duration of the flash is only in the order of 10^{-4} sec, thus much shorter than the flash of the "Photoflux" lamps and also much shorter than the exposures that could be reached with any form of mechanical shutter. In the case of these lamps, therefore, it is always the duration of the flash that determines the exposure time, and any "synchronization" amounts only to an automatic application of the "open-and-flash" method.

Something similar applies for the so-called paste lamps, which are small flashbulbs not containing any aluminium-magnesium wire and having only an explosive paste which of itself produces a fair amount of light, while the duration of the flash is extremely short and accordingly also the maximum time is short (5 - 10 msec). Some cameras, namely the Kodak Six-20 and the Kodak-Brownie Reflex, are specially designed for use with these lamps, being fitted with a synchronizer having a fixed and very short delay time; ipso facto they cannot be used with "Photoflux" flashbulbs.

Synchronization of flashbulb and focal-plane shutter

In the case of a focal-plane shutter two "curtains" with a slit between them pass across the photographic plate or film. This permits of very short exposures, e.g. 1/1200 sec. The total time taken for the slit to travel from one edge of the plate to the other, however, is much longer than the exposure at each point of the plate; a normal is 30 msec, but in some, large, cameras it may even be more than 40 msec.

For all parts of the plate to be uniformly exposed when taking a flash photograph with a focal-plane shutter, the luminous flux of the flashbulb would have to be constant during the whole of the running time of the shutter. Since this is not possible of realization it has been stipulated as a practical requirement that the minimum and the maximum

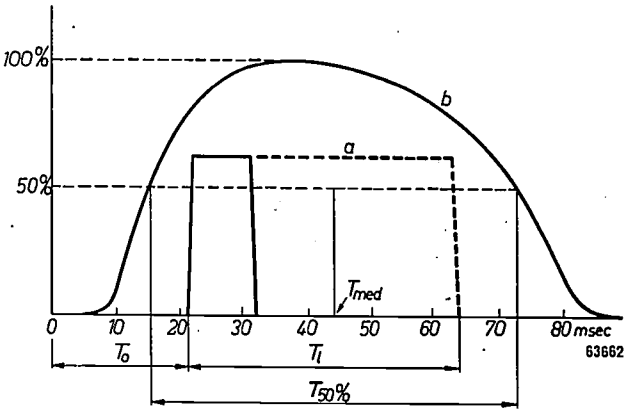


Fig. 9. *a*) Characteristic of a focal-plane shutter. The ordinate indicates the luminous flux passing through the slit. Owing to the movement of the slit however this luminous flux falls successively upon different parts of the sensitive plate: the exposure time of each part (indicated by the fully drawn lines for a point near the edge) is much shorter than the running time T_l of the slit.
b) Light-time characteristic of a flashbulb for use with a focal-plane shutter. The 50% flash time ($T_{50\%}$) has to be longer than the running time of the slit to avoid differences in exposure of the plate greater than a factor 2. The synchronizer has to start the shutter with a delay time $T_0 = T_{med} - \frac{1}{2}T_l$; T_{med} is the median time reckoned up to the middle of the 50% flash time.

[8]) The delay time should actually also be matched to the adjusted exposure time, but as this makes the matter very complicated one is usually satisfied if the ideal synchronization is obtained for only one exposure time. This must then be for the shortest exposure (say 1/500 sec), matching being the mos critical for this exposure.

[9]) See, e.g., S. L. de Bruin, An apparatus for stroboscopic observation, Philips Techn. Rev. 8, 25-32, 1946.

flux during that time shall in any case differ by no more than a factor of 2. From the light-time characteristic of a flashbulb it can be seen at once between what moments the lamp answers this requirement; these are in fact the moments at which the luminous flux has reached 50% of its maximum value, see *fig. 9*. The interval of time between these two points of the curve is called the "50% flash time". The time elapsing between the closing of the contact and the moment at which half of the 50% flash time has expired is denoted as the "median time".

For use with a focal-plane shutter one must therefore have a flashbulb with a 50% flash time at least equal to the running time of the shutter, while, independently of the exposure time, the synchronizer must have a delay time T_0 equal to the the median time less half the running time (cf fig. 9).

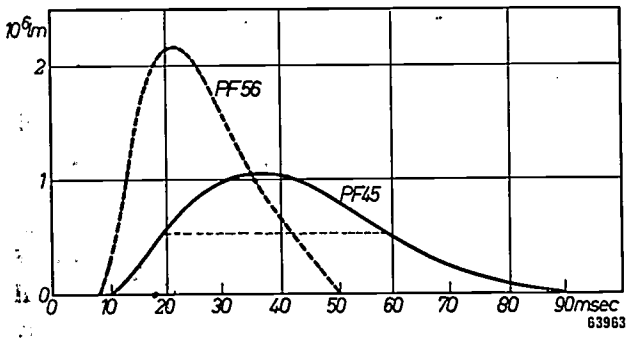


Fig. 10. Light-time characteristic of the "Photoflux" lamp PF 45, which has a 50% flash time of about 40 msec. (Note the differences of the scales of this figure and fig. 7!) For the sake of comparison the characteristic of the PF 56 has also been indicated (broken line).

It is for this purpose that the PF 45 has been developed. By a suitable choice of the gas filling the combustion of the Al-Mg wire in this type of "Photoflux" lamp is so delayed as to give the flat light-time characteristic reproduced in *fig. 10* [10]). The 50% flash time amounts to approximately 40 msec, which is long enough to allow of the lamp being used in combination with the majority of present focal-plane shutters, of small as well as large cameras.

It is of course also possible to use the PF 45 in combination with a between-the-lens shutter, but this cannot be said to be efficient. It has in fact only been possible to get the great width of the characteristic by accepting a relatively small maximum luminous flux, considering the very large total amount of light. With short exposures with an objective shutter a much greater effect is therefore obtained with the PF 56, which yields about the same total amount of light.

As already remarked, in appearance the PF 45 and the PF 56 are exactly alike, the difference only being seen from the type number stamped in the aluminium plate inside the neck of the bulb. To minimize the risk of mistakes being made, the type number of the PF 45 is marked in red letters.

To say that the PF 45 is not very efficient for short exposures with a between-the-lens shutter is only another way of giving expression to the fact that in the combination of a focal-plane shutter and a flashbulb the light that is available is rather inefficiently utilised. An improvement can only be made by shortening the running time of the slit. In the "Contax" camera, for instance, the running time has been reduced to about 15 msec by arranging for the slit to travel along the short side instead of the long side of the negative and, furthermore, by employing a stronger spring for driving the curtains. For this camera, therefore, the normal flashbulbs can be used, such as the PF 56 or the PF 25 (with 20 and 15 msec 50%-flash time respectively), with which for short exposures more economical use is made of the available light, owing to the high peak. On the other hand also the PF 45 can still be used, in which case another advantage of the short running time is to be derived, namely that, since only a part of the 50% flash time is used, the differences in luminous flux are much smaller than the permissible factor 2.

Summary. A description is given of the construction of the "Photoflux" flashbulbs. The light is produced by combustion of a ball of aluminium-magnesium wire in oxygen. An internal and an external coating of lacquer help the bulb to withstand the impact of the pressure resulting from the combustion. Up to now types of "Photoflux" lamps are being made, the PF 110, PF 56, PF 45, PF 25 and PF 14. The number behind the letters PF denotes the total amount of light in thousands of photographic lumen-seconds. The PF 56, PF 25 and PF 14 reach their maximum output of light about 22 msec after the electric circuit for the ignition is closed, so that they can be used in combination with practically any of the present synchronizers for between-the-lens shutters. The PF 45, which has a flatter light-time characteristic (the 50% flash time is about 40 msec), has been specially developed for use with cameras having focal-plane shutters.

[10]) Other ways of changing the light-time characteristic in this sense are: eccentric ignition of the ball of wire, or the use of two kinds of wire (or a combination of wire and foil) in the bulb. In the latter case, owing to one of the components burning earlier and quicker, we get the sum of two characteristics with mutually displaced peaks, which, if the "saddle" between the peaks is not too deep, yields a characteristic similar to that given in fig. 10.

AN INSTRUMENT FOR RECORDING THE FREQUENCY DRIFT OF AN OSCILLATOR

by W. W. BOELENS.

621.396.615.12:621.317.76

A quartz oscillator has a frequency that is sufficiently constant for determining the frequency drift of an ordinary valve oscillator due to changes in temperature, for which purpose it is in fact the most suitable oscillator. This frequency drift, for instance that of an oscillator in a superheterodyne receiver, should be measured, however, in more than one position of the tuning capacitor. The problem is how to derive from one quartz crystal a number of reference frequencies spread more or less regularly over the frequency range of the oscillator to be measured. With the frequency-drift meter described here this problem has been solved by generating in a number of frequency-multiplying stages the twenty-fourth multiple of the crystal frequency and by an artifice deriving therefrom nine other multiples (the 19th up to and including the 28th).

It is a well-known phenomenon that after a valve oscillator has been switched on its frequency tends to drift. This is due to the fact that when an oscillator is switched on various sources of heat are brought into play (valves, resistors, transformers), so that the temperature and thus also the dimensions and hence the self-inductance or capacitance of the elements in the oscillating circuit undergo a gradual change. In practice it often takes some hours before a nearly stable condition is reached.

If the oscillator in question is part of a superheterodyne receiver and no adequate steps have been taken to counteract this phenomenon then, as a result of this frequency drift, the tuning of the apparatus may have to be readjusted several times to ensure satisfactory reception of a particular station. This is especially important in the reception of waves shorter than 10 m. Steps taken to prevent this may consist in the application of thermal insulators around the sensitive parts, the provision of ventilation, the use of circuit elements having a low temperature coefficient, etc.

Thus, after having applied one or more of these measures in an experimental apparatus, the designer of a receiving set is faced with the task of ascertaining whether the defect in question has been sufficiently remedied. This he does by measuring, with reference to a reference frequency, the frequency drift Δf_0 as a function of the time t after switching on, and this involves a series of measurements that may have to be extended over a number of hours. As a rule the first result will not be satisfactory; a suspected element may have to be replaced and the measurements then repeated all over again.

It does not suffice simply to measure the final value of the drift, because it may have previously reached much higher values. Thus the drift (and by that we mean here the absolute value of the frequency variation) as a function of time may show a maximum somewhere on the curve (*fig. 1*). This is due to the fact that the oscillating circuit comprises a number of elements (inductance of the coil and of the wiring, capacitance of the tuning capacitor, of the wiring and of the valve) some of

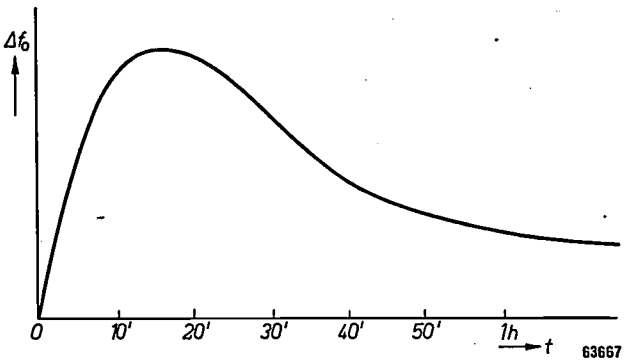


Fig. 1. Example of the variation of the frequency drift Δf_0 of an oscillator as a function of the time t after switching on. In the case drawn there is a maximum.

which may show a positive and others a negative temperature coefficient of the inductance or of the capacitance as the case may be. In course of time there may be a certain degree of compensation which does not exist at the beginning. It cannot be predicted at what stage the maximum may occur after switching on because of the great differences in the rate at which the various components reach their final temperature: the valves and those parts soldered onto the valve sockets may attain their final temperature in 10 or 15 minutes, whereas the parts mounted on the chassis may take a much longer time.

Consequently it is necessary to take the measure-

ments at sufficiently short intervals of time to ensure that any maximum occurring in the frequency drift does not escape notice. First they should be taken, for instance, every thirty seconds, then every minute, after a quarter of an hour every five minutes and after an hour every fifteen minutes. These intervals are not long enough to permit the observer to occupy himself seriously with any other work in between, so that for that reason alone such drift measurements cost a great deal of time.

Furthermore, the plotting of one drift curve alone does not suffice, it being necessary to plot a curve for several positions of the tuning capacitor, since the various capacitances in the oscillating circuit vary according to different functions of time, so that measurements have to be taken for different ratios of the component capacitances.

Very much time can therefore be saved by employing a recording instrument which, once it is started, automatically plots the drift curve. Below a description will be given of such a recording drift meter in a form specially designed for testing F.M. radio receivers, the effective frequency range of which extends from 88 to 108 Mc/s.

Principle of the drift meter

F.M. receivers usually work with an intermediate frequency of 10.7 Mc/s, though lower values may also occur. It depends upon local conditions (near-by transmitters working on adjacent frequency channels) whether the designer finds it better to choose an oscillator frequency f_o for the receiver that is higher or lower than the frequency of the stations to be received, with the result that in practice both solutions occur. Therefore, in order to be of use for as many types of receivers as possible, the drift meter must be designed so as to cover a range extending from $88-10.7 = 77.3$ Mc/s to $108 + 10.7 = 118.7$ Mc/s.

The maximum drift that can be tolerated in this frequency range varies from 25 to 50 kc/s, according to the quality of the receiver, and thus the drift does not amount to more than about 0.05%, which is far too small to be measured directly with sufficient accuracy. The beat frequency of f_o is therefore formed with a neighbouring reference frequency f_r; this beat frequency shows a relatively much larger and thus more easily measurable drift.

The reference frequency f_r cannot be taken from a valve oscillator with an ordinary oscillating circuit because f_r would then itself show variations of the same order as the drift of f_o. For generating the reference frequency f_r one must therefore use a crystal-controlled oscillator. The difficulty that a crystal gives only one fixed value of f_r, whereas f_o is to be measured for a number of values (in the range from 77.3 to 118.7 Mc/s), can be met by choosing a crystal frequency f_c much lower than f_o, say $f_c \approx 4$ Mc/s, and producing, by n-fold frequency multiplication, a sufficient number of multiples of f_c which lie within this range and thus can serve as reference frequencies f_r. With $f_c = 4.232$ Mc/s (the value actually used) ten useful reference frequencies are obtained: $f_r \approx 80.4$, 84.6, 88.9 ... 118.5 Mc/s, a sufficient number of which will always lie in the range covered by the oscillator, no matter whether the oscillator frequency is higher or lower than the signal frequency, and for all intermediate frequencies up to well over 10 Mc/s. To get these ten reference frequencies it is necessary that n can assume the values 19, 20, 21 ... 28.

Now this is easier said than done. In the frequency range specified a 20-fold multiplication with such a selectivity that the adjacent multiples do not interfere cannot be brought about in one single multiplying stage [1]). It is true that good results could be reached by employing more than one stage, each with a small multiplication factor, but of course this only helps for those values of n which can be resolved into small factors, such as $24 = 4 \times 3 \times 2$, and $27 = 3 \times 3 \times 3$, so that the problem of getting $n = 19$, 22, 23, etc. cannot be solved in this way. To overcome this difficulty the following artifice has been employed.

As indicated in the block diagram of *fig. 2*, from the crystal frequency $f_c = 4.232$ Mc/s a 24 times as high reference frequency of 101.57 Mc/s is obtained in three stages. Both a voltage of this frequency $24 f_c$ and the normal voltage of the oscillator frequency f_o are applied to the mixing valve M_1 of the receiver under test. The voltage at the output of this valve has the much lower beat frequency $|24 f_c - f_o|$. After amplification, accompanied by the removal of undesired combination frequencies by means of a tuned circuit, this output voltage is applied to the input *1* of a second mixing valve M_2 forming part of the drift meter. To the input *2* of the latter valve a voltage is applied having the crystal frequency f_c itself and of such an amplitude and bias that the valve is rendered conductive

[1]) In principle the series of harmonics obtained without this selectivity are just suitable for the purpose, but the components are so small in amplitude that considerable amplification is necessary, and at frequencies of about 100 Mc/s this gives rise to great difficulties. It is furthermore unknown which of the harmonics is serving as a reference frequency.

only for a short time at the positive peaks of the voltage at the input *2*. This has the same effect as if this were a pulsatory voltage with the repetition frequency f_c. Consequently the valve shows a conversion conductance — the ratio of the A.C. output current with the beat frequency to the A.C. input voltage at *1* — which, as a function of the

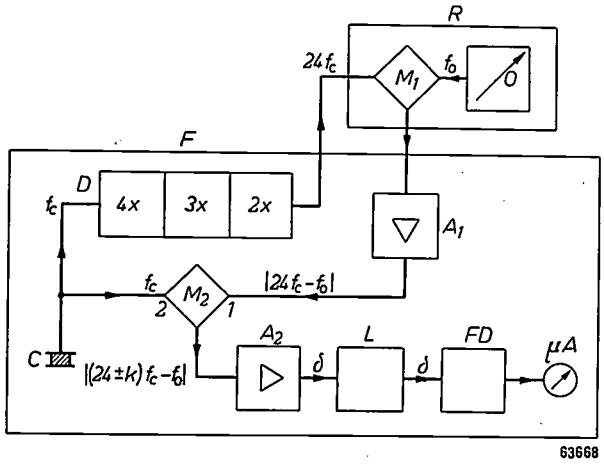


Fig. 2. Block diagram showing the principle of the drift meter. *R* receiver with oscillator *0* (frequency f_o) the drift curve of which is to be recorded, and mixing valve M_1. *C* quartz oscillator (frequency f_c), *D* frequency multipliers (24 ×), M_2 second mixing valve, A_1 and A_2 amplifiers, *L* limiter, *FD* frequency detector, μA recording moving-coil meter with scale calibrated in frequencies. The parts within the rectangle *F* belong to the drift meter.

frequency at input *1*, is approximately equal for frequencies f_c and not too large multiples of f_c. Thus, in the main, the output voltage of the valve contains not only components with the frequencies $|24\ f_c - f_o \pm f_c|$ but also components with $|24\ f_c - f_o \pm k\, f_c| = |(24 \pm k)\ f_c - f_o|$, where $k = 1, 2, 3$ and so on. The shape of the pulse is such that the components with $k = 1 \ldots 5$ are present with the required amplitude. The effect is therefore as if the reference frequencies $19f_c \ldots 23f_c$ and $25f_c \ldots 28f_c$ were all present simultaneously (the omitted term $24f_c$ will be dealt with later; $29f_c$ is not in the desired range and can be left out of consideration). After f_o has been brought sufficiently near one of the above reference frequencies by turning the tuning knob of the receiver, the small difference δ between these two can easily be separated with a low-pass filter from the much larger differences existing between f_o and the other multiples of f_c and which are about equal to f_c or several times f_c (thus 4 Mc/s or higher). If the frequency f_o of the oscillator under test changes by an amount Δf_o, then δ changes by the same amount, and this amount is usually considerable compared with δ. How this drift is detected and recorded will be shown below.

Let us now consider the case where f_o lies in the neighbourhood of $24f_c$, and let the difference again be δ. This difference will then be present already at the output of the first mixing valve, so that in this particular case there is no need of the second mixing valve. Thus the second mixing valve could in this case be by-passed, but it has been found that the switches which would then be required can be dispensed with and the circuit left unchanged, because in spite of the valve being unblocked and blocked with the high frequency f_c it still transmits the low frequency δ with sufficient amplitude.

By means of the second mixing process described — the mixing of $|24\ f_c - f_o|$ with f_c — it has thus been found possible to transpose the drift to a much lower frequency range, with only one fixed multiplication factor (24) of the crystal frequency, for ten different frequencies.

Further, it is necessary to record the relatively low output frequency δ of the second mixing valve varying according to the amount of the drift. This can be done, for instance, by generating a direct current which is (approximately) proportional to the frequency δ and causing this current to flow through a recording meter.

To produce such a direct current we have made use of a known frequency detector circuit. This is preceded by an amplifier (A_2 in fig. 2) and a limiter. The amplifier is needed because the output voltage of the second mixing valve is too small for the frequency detector to function properly; moreover it contains the filters blocking the voltage components with undesired high frequencies. The limiter clips the peaks of the amplifier output, so that an almost square wave voltage of constant amplitude is produced. The latter is necessary because the output current of the frequency detector has to be independent of the amplitude of the output voltage from the second mixing valve.

From the fact that the meter records only the absolute value of the difference between one of the standard frequencies $n f_c$ and the oscillator frequency f_o it follows that the recorded curve (e.g. *fig. 3a*) leaves one in doubt as to whether f_o was smaller than nf_c (fig. 3*b*) or greater (fig. 3*c*). This question can be solved by slightly turning the tuning knob of the receiver such as to make f_o somewhat larger and then noting in which direction the deflection of the meter is thereby changed; if the meter reading is reduced then f_o was smaller than nf_c; if it is increased f_o was greater.

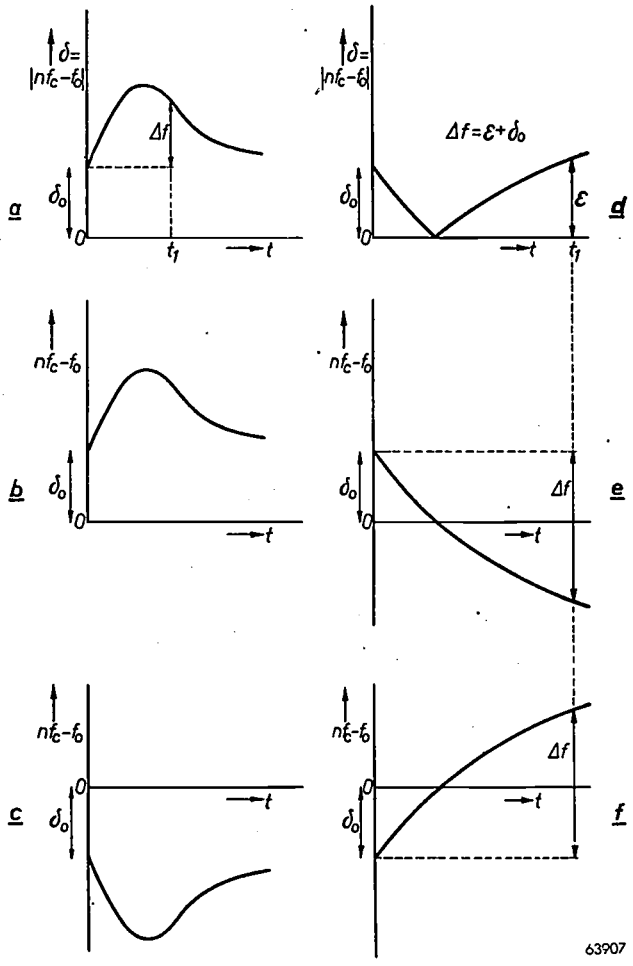


Fig. 3. *a*) Example of a recording. The frequency drift Δf at the moment $t = t_1$ is the reading at t_1 less the original deflection δ_0. The variation of $nf_c - f_o$ may have been according to (*b*) or (*c*); it is easily determined which of these two possibilities it was by purposely increasing f_o a little.
A turning point in the recording (*d*) is to be ascribed to $nf_c - f_o$ passing through zero according to (*e*) or according to (*f*). In (*d*) the drift at the moment $t = t_1$ is obtained by adding the initial value δ_0 to the deflection ε.

The recorded curve may also show a turning point on the zero line (fig. 3*d*), this being the case when the difference between nf_c and f_o has passed through zero during the test. The trend of $nf_c - f_o$ may then have been according to fig. 3*e* or fig. 3*f*. By purposely increasing f_o a little it can be decided which of the two applies.

The manner in which the amplitude of the frequency drift has to be read from the recorded curve is indicated in figs 3*a* and 3*d*.

Form of execution

We shall now briefly discuss some details of the various component parts.

Crystal oscillator

Owing to use being made of a quartz crystal with a "temperature-independent" cut, the frequency f_c does not change by more than $1 : 10^6$ per °C temperature variation. This is so little that in our case it is not necessary to place the crystal in a thermostat, and this is fortunate in that it avoids the long waiting time involved in thermally balancing the thermostat. By providing ample ventilation openings in the casing containing the apparatus and also protecting the crystal with heat-insulating material, it has been possible to limit the final value of the frequency drift of the crystal to about 40 c/s and thus the drift of the standard frequency to about 1000 c/s, which is sufficiently small for our purpose.

The functioning of the crystal oscillator and the multiplying stages can be checked by measuring the grid current of the valves employed in these stages, for which purpose a micro-ammeter with switch is provided.

Amplifiers and mixing stage

In *fig. 4* a somewhat simplified diagram is given representing the circuit of the first amplifier (A_1 in fig. 2, with an EF 40 valve) and the second mixing valve M_2 (the hexode part of an ECH 41 valve). Between these two valves is a circuit, which can be tuned to the frequencies f_c, $2f_c$, $3f_c$ $4f_c$ and $5f_c$, according to whether the drift is to be measured at $f_o \approx 23f_c$ or $25f_c$, $22f_c$ or $26f_c$, ... $20f_c$ or $28f_c$, or $19f_c$ respectively. In these five cases the absolute bandwidth has to remain the same, this being achieved by switching only the inductance and leaving the capacitance and parallel resistances unchanged. The circuit is so arranged that the gain at a frequency deviating 0.2 Mc/s from the resonance frequency (the drift can be measured to 0.1 Mc/s maximum) is not more than 3 dB less than at resonance. For measuring at

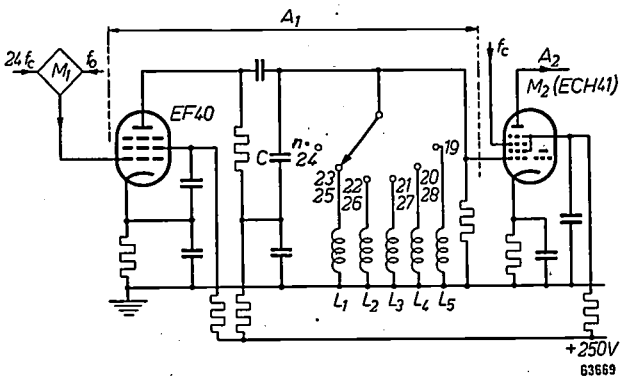


Fig. 4. Amplifier A_1 of fig. 2, with the mixing valve M_1 of the receiver and the mixing valve M_2 belonging to the drift meter. Between the pentode EF 40 and the hexode ECH 41 is a tuned circuit formed by the capacitor C and one of the coils L_1 ... L_5 according to the desired value of n; for $n = 24$ the inductance branch is interrupted.

$f_o \approx 24 f_c$ the inductance branch is interrupted, so that an ordinary resistance-capacitance coupling is left.

Limiter and frequency detector

The circuit of the limiter and of the frequency detector is represented in *fig. 5*.

The limiter works with two diodes (D_1, D_2). The cathode of D_1 is kept at a constant potential $E \approx +20$ V.

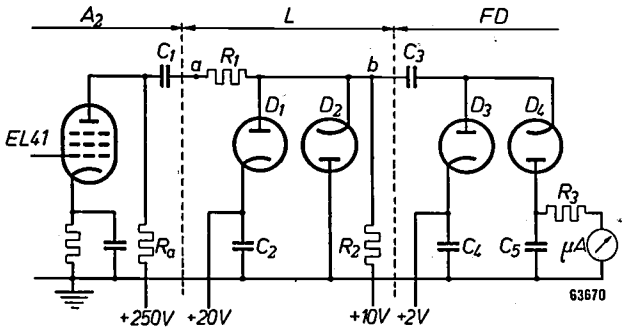


Fig. 5. Circuit diagram of the limiter (L) and of the frequency detector (FD). EL 41 output valve of the amplifier A_2 (fig. 2), with the anode resistor R_a. D_1 ... D_4 diodes EA 50. C_1, C_3 coupling capacitors, C_2, C_4, C_5 smoothing capacitors, μA micro-ammeter with resistor R_3. The function of the resistor R_2 is to keep the mean potential of the point b at + 10 V.

At the point a there is practically the same alternating voltage as at the anode of the output valve in the amplifier A_2; the amplitude of this alternating voltage is much greater than E. The diode D_1 passes current only when the alternating voltage at a is greater than E, and the diode D_2 conducts when the potential of a is negative. Disregarding voltage loss in the diodes, the potential of the point b is then $+E$ and zero respectively. The transition between these two values takes place in the very short intervals during which no current flows through either of the diodes.

Thus the limiter supplies an almost square-wave output voltage, with alternately the values $+E$ and zero. Since b is connected via the resistor R_2 to a point with the potential $\frac{1}{2}E$, the positive peaks of the alternating voltage at a are cut off as much as the negative peaks, so that the intervals during which the voltage at b assumes the values $+E$ and zero have the same length.

The frequency detector, which is of a known type [2]), similarly has two diodes (D_3, D_4). Connected in series with D_4 are a resistor R_3 and a micro-ammeter, these being shunted by a smoothing capacitor C_5. In order to avoid the possibility of the diodes causing a certain zero current to flow through the meter while no external voltage is applied, the cathode of D_3 is biased with a small positive potential $E_0 \approx 2$ V.

The formula for the current I flowing through the meter is:

$$I = \frac{\delta C_3}{1 + \delta C_3 R_3} (E - E_0), \quad \ldots \quad (1)$$

where δ is the frequency of the pulsating voltage at the terminal b, C_3 the value of the capacitor connected in series with the diodes, while R_3 includes the resistance of the micro-ammeter.

The formula (1) can be derived in the following way, ignoring the voltage loss in the diodes.

When the voltage at the terminal b is zero, in the stationary state there is a voltage across C_3 equal to the voltage $V = IR_3$ across C_5. When the voltage at b suddenly changes from zero to $+E$ then C_3 receives via the diode D_3 a charge equal to $C_3 (E - V - E_0)$. When the voltage at b drops again from $+E$ to zero this charge flows via the diode D_4 to R_3 and C_5 connected in parallel. This takes place δ times per second, so that the mean current I amounts to

$$I = \delta C_3 (E - IR_3 - E_0).$$

Solving this equation for I leads to formula (1).

As can be seen from formula (1), the direct current I is practically proportional to the frequency δ if care is taken that $\delta C_3 R_3 \ll 1$. This is what is in fact done in cases where the detection has to be as linear as possible. In our case, however, the meter need not record the frequency on a purely linear scale; in fact this is even undesirable, because with high values of δ, which occur while tuning f_o near to one of the reference frequencies, the meter would be overloaded. For this reason R_3 and C_3

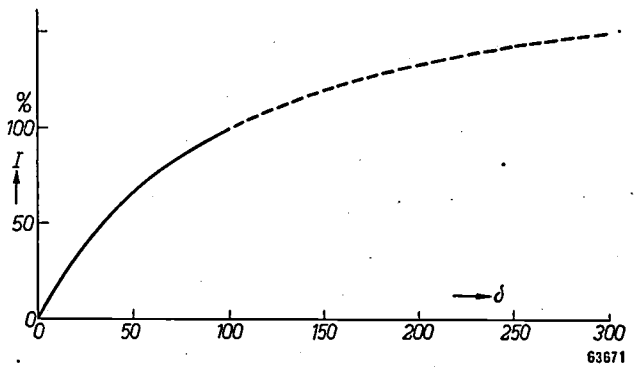


Fig. 6. The direct current I flowing through the micro-ammeter (in % of the current required for full deflection) as a function of the frequency δ in kc/s at the input of the frequency detector.

[2]) S. W. Seeley, Ch. N. Kimball and A. A. Barco, Generation and detection of frequency-modulated waves, R.C.A. Review 6, 296-286, 1942; reprinted in Frequency Modulation, Vol. I, pp. 147 et seq. (R.C.A., Princeton, N. J., U.S.A.).

Fig. 7. Frequency-drift meter, in a form suitable for mounting in a rack. The left-hand meter, calibrated in kc/s, is used for preparatory adjustments; the recording meter is not shown in the illustration. The right-hand meter is used to check the various grid currents selected by the knob in between them. Underneath is the knob for selecting the desired reference frequency (n = 19 ... 28), cf. fig. 4.

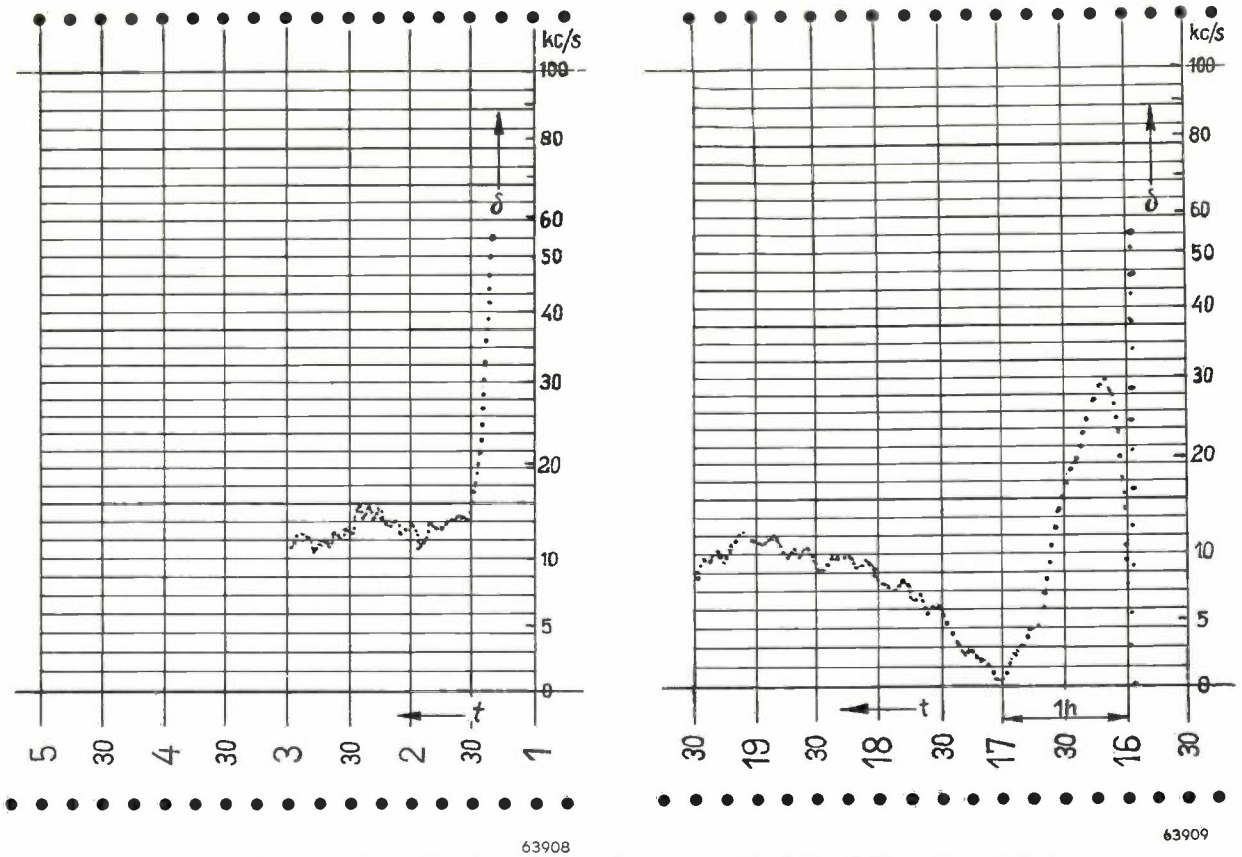


Fig. 8. Two recordings. The first shows a just acceptable drift of 55 — 12 = 43 kc/s. The second recording shows a turning point five minutes after the start, and at the end of 15 minutes a maximum at which the total drift is 55 + 30 = 85 kc/s. After a second turning point a second maximum is reached, at about three hours from the start; it shows a frequency drift of 30 + 10 = 40 kc/s with respect to the first maximum. If only the difference between the frequencies at the start and that at a moment three hours later had been measured the acceptable amount of 55—10=45 kc/s would have been found, but the quite unacceptable maximum drift of 85 kc/s would have remained unnoticed.

have been so chosen that with $\delta = 100$ kc/s the value of $\delta C_3 R_3 = 1$. The curve $I = f(\delta)$ then assumes the shape represented in *fig. 6*. There thus is no longer any risk of overloading, while the frequency is still sufficiently legible over the whole range, especially so with small values of δ, so that even a small frequency drift can still be accurately determined.

In the form in which it is shown in *fig. 7* the drift meter has two micro-ammeters connected in series with the resistor R_3 (fig. 5): a normal built-in meter and a recording instrument (full deflection at 34 μA) kept separate. While the test is being started the readings are taken from the built-in meter. The frequency scale of the built-in meter runs up to 150 kc/s, and that of the recording meter to 100 kc/s.

Finally, in *fig. 8* two recordings are reproduced, one showing an acceptable drift and the other an intolerably large drift.

Summary. The frequency-drift meter described here has been designed for measuring the frequency variation of the oscillator in various types of superheterodyne F.M. receivers in the 88-108 Mc/s band. The drift meter is controlled by a quartz crystal with frequency $f_c = 4.232$ Mc/s, of which the 24-fold frequency (101.6 Mc/s) is obtained with the aid of multiplying stages. In the mixing valve of the receiver this harmonic is mixed in the usual way with the oscillator voltage having the frequency f_o that is to be observed, The output voltage, which contains, inter alia, the beat frequency $|24 f_c - f_o|$, is mixed with a voltage of the crystal frequency in a second mixing valve incorporated in the drift meter and so adjusted that it is rendered conductive only at the peaks of the latter voltage. The result is that from this second mixing process frequencies such as $|24 f_c - f_o \pm k f_c| = |(24 \pm k) f_c - f_o|$, with $k = 1, 2...$ are obtained, so that, with only one multiplier working with good selectivity, a series of reference frequencies, $(24 \pm k) f_c$, are available which can be separated by means of a circuit tuned to one of the frequencies $k f_c$. In this way, in the drift meter described ten reference frequencies are available, from $19 f_c = 80.4$ Mc/s up to and including $28 f_c = 118.5$ Mc/s.

After amplification and limitation, the output voltage of the second mixing valve, with frequency $|(24 \pm k) f_c - f_o|$, produces at the output of a normal frequency detector a direct current which is approximately proportional to the frequency variation and thus also to the drift of f_o. This direct current is recorded as a function of time by a recording meter and thus a great deal of time in the regular checking of the drift of receivers is saved.

THE PERMISSIBLE BRIGHTNESS OF LAMP FITTINGS

by D. VERMEULEN and J. B. DE BOER. 628.95:628.972:535.241.44

The interest in the problems of "brightness engineering" is to a certain extent typical for changes that have taken place in the last decades in our conceptions about the working man. Whereas formerly interest was devoted mainly to the physiological conditions for the proper performance of a task, nowadays also the psychological factors that play a part are considered, and where formerly one spoke of efficiency now one speaks of "comfort".

Brightness engineering

With the development of the sources of light in the last 50 years it has become possible to reach high levels of indoor illumination in an economical manner. For the period up to about 1935-1940 the use that has been made of this possibility may be described as the adequate illumination of the working plane. It was realized that for numerous tasks to be performed properly, both in industry and in the home, reasonably high levels of illumination were needed, and from the experience gained in this field lighting standards were drawn up, in the form of tables giving the recommended illumination for all possible kinds of work [1]).

The principle underlying such standards is in all cases: it must be possible to see clearly and quickly what has to be seen when performing a task. However, as lighting equipment came to be installed more and more in accordance with this principle it became evident that it is not sufficient to consider only the visibility of the object upon which one is working. While a task is being performed the eyes are not directed continuously upon the work in hand, there being always shorter or longer intervals during which one looks up from one's work into the surroundings in the normal way. These intervals should be a rest to the eyes, a brief relaxation from concentration upon the work: thus the lighting should give a feeling of comfort.

It is particularly with the high levels of illumination now often applied that this additional requirement assumes importance. The condition of "comfortable" lighting implies that the surfaces upon which the eye falls when looking up must not contrast too strongly in brightness with the work area. Consequently, on the one hand there must not be any very dark corners, while on the other hand the greatest brightness in the surroundings (i.e., in the case of direct lighting, the brightness of the lamp fittings) must not be so great that one becomes unpleasantly conscious of its presence, or at least there must not be so many fittings in the room that the eye cannot very well escape them when looking up. It is just with high levels of illumination however that very many and/or very bright fittings have to be used, so that there is a great chance of the lighting causing discomfort to the eye [2]).

Since this new point of view concerning the distribution of brightnesses in the surroundings has come to the fore, in the course of the last 10-15 years illumination engineering has entered a new phase, the phase of what has been named in America "brightness engineering".

There are, of course, other factors which help to decide whether people working in a room for a long time find the lighting comfortable or not. We have only to mention the colours, shadows, possible glitter and the aesthetic appearance of the room. Here we shall leave these factors out of consideration and confine our attention to the question of brightnesses, and in particular to the brightness of the lamp fittings.

What the problem is, and previous attempts to solve it

The question to be put, therefore, is: what is the maximum permissible brightness of the lamp fittings for a person inside the room not to get a feeling of discomfort from the lighting system? To give an answer to this question that will satisfy practical needs it has to be dealt with in two stages. First a measure has to be determined for the border between "comfortable" and "uncomfortable" lighting. Secondly we must try to find some law

[1]) See, e.g., A. A. Kruithof and A. M. Kruithof, Basic principles for the formulation of illumination standards, Philips Techn. Rev. **10**, 214-220 1949, (No. 7).

[2]) W. Harrison, What is wrong with our fifty-footcandle installations? Ill. Eng. **31**, 208, 1937.

according to which it can be predicted whether certain conditions of illumination create a situation on this or that side of the limit. Such a law must of course be based upon the results of experiments (in these experiments the choice of the criterion is involved).

Among the investigations so far carried out in this direction we mention those undertaken by Luckiesh and Guth [3]) and by Harrison [4]). The first-mentioned investigators followed up the 20-years older measurements of Holladay [5]), which concerned glare and were obtained by what is known as the shock method. Holladay let a glaring source of light appear for a moment in the visual field of an observer. When the brightness of the light source is very high the observer's visual capacity is temporarily reduced, and this reduction can be measured. When the brightness of the light source is moderate there is no question of such a reduction of the visual capacity but the observer gets a feeling of discomfort. The degree of this "discomfort glare" was determined by Holladay, and later by Luckiesh and Guth, by asking the observer to judge — immediately after the appearance and subsequent disappearance of the glaring light — in what degree, from scarcely noticeable to quite intolerable, he would place the feeling of discomfort. In this way the discomfort glare was determined as a function of a series of variables: the brightness B_a and the dimensions of the light source, the brightness B_s of the surroundings, the angle α between the direction of vision and the direction from the eye to the light source, and the horizontal distance d away from the lamp.

Harrison worked out all this experimental material into a formula. To each source of light in a given lighting situation he ascribed a glare factor (f) as a measure for the discomfort glare caused by that light source. For this glare factor he gave the formula (somewhat simplified by us):

$$f = k \frac{B_a^2 \omega}{B_s^{0,6} \sin^2 \alpha}, \quad \ldots\ldots \quad (1)$$

where ω is the solid angle under which the observer sees the light source and k is a factor the further definition of which is of no significance here. When there are several light sources together in the field of vision — so Harrison assumed — the various facors f could be added up. The assumption that these factors f would be additive can be made plausible by the consideration that f can be interpreted as the surface of a source of light (with fixed brightness) necessary in a standard situation to cause the same degree of discomfort glare as that caused by the light source in question.

For the total glare factor Harrison gave as criterion that, for the lighting not to cause any feeling of discomfort, this must not exceed 15.

This manner of representation has its value as a first attempt to arrive at a quantitative treatment of the problem, but it has to be pointed out that in the experimental material underlying this manner of approach there are some assumptions which make its applicability to our problem seem doubtful.

In the first place let us consider the criterion. The situation in which the observer finds himself placed for experiments according to the shock method is entirely different from that which we have in mind when considering whether a lighting system is comfortable or not: it is a matter of the feeling of discomfort that may be experienced when staying for a long time in a lighted room.

In the second place, the experiments were carried out only with round or practically round sources of light. Considering the extent to which tubular fluorescent lamps are now being used, especially for high levels of illumination in large rooms and halls, obviously it is questionable whether the results reached by Holladay and by Luckiesh and Guth, with Harrison's conversion to equivalent surfaces (glare factors), can be taken for granted as being valid also for elongated sources of light.

These and other considerations have led the Illumination-Engineering Laboratory of Philips Works in Eindhoven to carry out a series of new experiments with as starting point the same question as put at the beginning of this section of our present paper [6]).

Set-up and manner of conducting the experiments

In our experiments we aimed at making the criterion approximate as closely as possible to actual practical conditions. From the ceiling of the room a lamp fitting was suspended whose brightness could be controlled by the observer himself. The lamps used for the general lighting

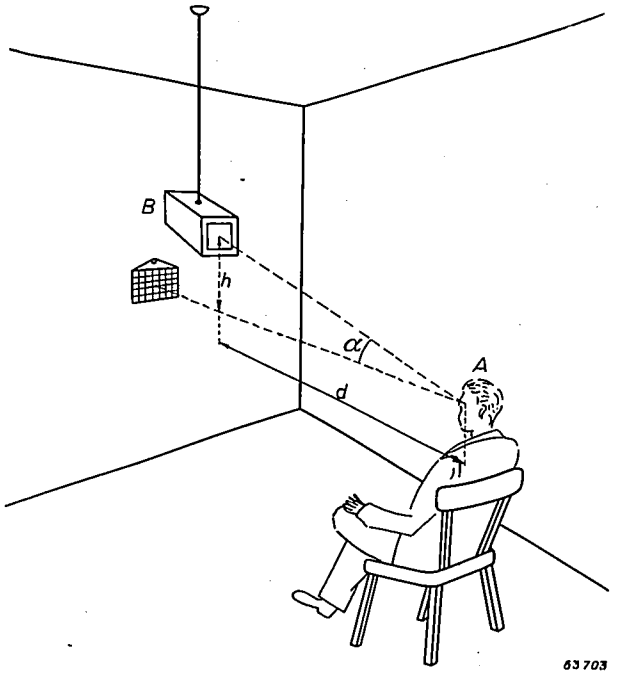
[3]) M. Luckiesh and S. K. Guth, Discomfort glare and angular distance of glare sources, Ill. Eng. **41**, 485, 1946.

[4]) W. Harrison, Glare ratings, Ill. Eng. **40**, 525, 1946; W. Harrison and P. Meaker, Further data on glare ratings, Ill. Eng. **42**, 153, 1947.

[5]) L. L. Holladay, The fundamentals of glare and visibility, J. Opt. Soc. Amer. **16**, 271, 1926.

[6]) These experiments have already been reported in a paper prepared by D. Vermeulen and J. B. de Boer for the conference of the International Commision on Illumination held in Paris in 1948; this report will shortly be published in Applied Scientific Research.

of the room were invisible to the observer and could be adjusted independently of the test light source. The observer was asked to imagine himself as being seated in an office and engaged in his normal work. Then, looking up in a horizontal direction,



63703

Fig. 1. Set-up for determining the permissible brightness of a square lamp fitting. The observer (A) himself adjusts the brightness of the light box (B) to what he judges to be just a tolerable value. The ambient brightness can be varied independently of the light box, as also the suspension height h of the box and the distance d from the observer. The luminous plane observed is always perpendicular to the viewing direction (thus it remains square). The tests were carried out with a number of light boxes of different sizes.

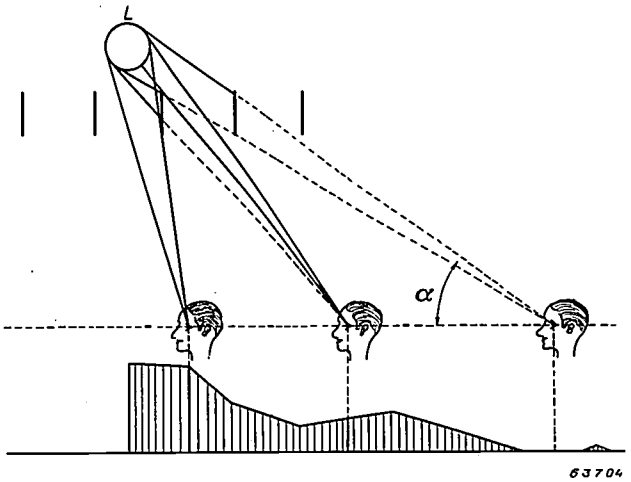
he had to adjust the brightness of the test fitting suspended in his field of vision until the limit of what he felt to be the tolerable brightness was reached.

Of course this limit is not sharply defined, the tolerability of a brightness observed in this way being judged from personal impression, so that great differences in the results are to be expected. Nevertheless it will presently be shown that with this, in itself vague, criterion a fairly well determined limit can indeed be found.

For the first series of tests we used as lamp fitting a square light box which can be taken as being the equivalent of a diffuser such as is commonly used for office lighting with incandescent lamps. Boxes of different sizes were suspended at various heights from the ceiling, so that the solid angle ω and the viewing angle α under which the fitting was observed could be varied independently of the likewise variable horizontal distance d from the observer to the fitting (*fig. 1*). Together with the ambient brightness already mentioned there were thus four independent variables as a function of which the brightness adjustments of each observer had to be determined. In this way data were obtained with which, taking our new criterion as basis, a law could be formulated analogous to that of formula (1).

In a second series of experiments the square light box of fig. 1 was replaced by an oblong one, by way of imitation of a fitting in which tubular fluorescent lamps could be used. Since the question in this case was mainly to decide whether a representation like that given by Harrison (conversion to equivalent surfaces) would hold for light sources of any shape, this series of tests was on a less extensive scale, only one light box being used, of certain dimensions (1.40 m long, 0.10 m high, placed horizontally, with the luminous surface always perpendicular to the line of sight).

Finally another series of tests were carried out with a somewhat different object, particularly concerning the construction of the well-known fittings with so-called louvres for fluorescent lamps. Practice has shown beyond doubt that the relatively low brightness of the fluorescent lamp (about 0.4 stilb) is still not low enough for these lamps to be mounted without any screening in the field of vision; the farther the lamps are removed from the central part of the field of vision (the larger the viewing angle α), however, the less trouble is experienced from the direct radiation of the lamps.



63704

Fig. 2. Schematic representation of a lamp fitting with louvres for fluorescent lamps. Owing to the partitions the observer sees less of the lamp the farther he stands away from it and, thus, the smaller the viewing angle α. The ordinates of the hatched plane indicate the visible surface of the lamp at each distance.

A similar diagram can be drawn for the case where the lamp lies in the plane of the drawing and not, as here, perpendicular to that plane. Since "TL" lamps, thus viewed in the longitudinal direction appear to have a lower brightness than when viewed transversely, the partitions are allowed to be more widely spaced in this direction.

By placing vertical screens under the lamp(s) the surface visible to an observer is reduced, the more so the farther the observer is away from it, and thus the smaller the angle α with respect to the observer when he is looking in a horizontal direction (*fig. 2*). In this situation, therefore, it is not asked what is the permissible brightness with a given surface, but inversely what is the permissible surface with a given brightness. Our set-up was accordingly adapted by confronting the observer with a horizontally suspended, naked, fluorescent lamp, the visible surface of which he was able to vary with the aid of a horizontal slit of variable width (*fig. 3*). He was then asked

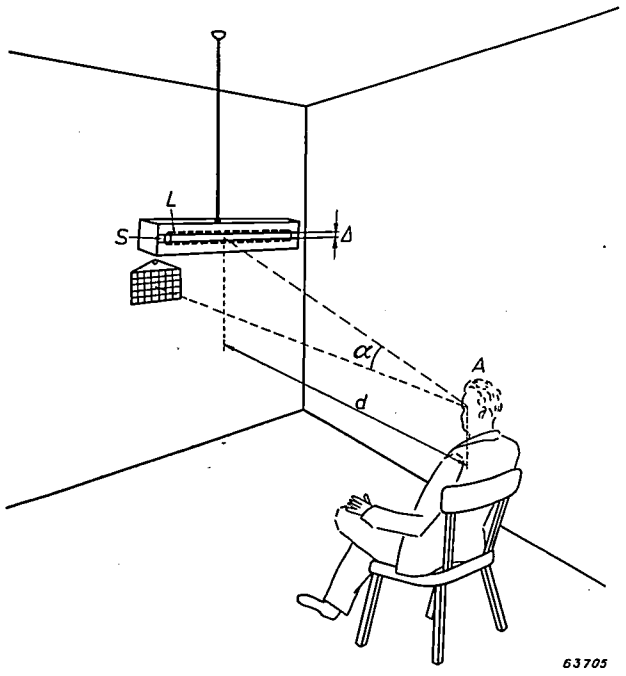


Fig. 3. The width Δ of the slit S through which the naked fluorescent lamp L is seen can be adjusted by the observer (A) himself to the value he judges to be just tolerable.

to adjust the width of the slit Δ (height of the visible lamp surface), while looking in the horizontal direction, so that the presence of the lamp did not cause him to feel any discomfort. These experiments were carried out with lamp lengths of 0.60, 1.20 and 2.40 m, to imitate the case where a number of fluorescent lamps are mounted in a line.

It has already been pointed out that the adjustments obtained in one and the same situation must inevitably show a great spread. An obvious method to cope with these difficulties is to define the required permissible brightness (or width of slit) as that brightness which 50% of the observers regard as being just tolerable. When plotting for each brightness value the percentage of observers whose adjustments exceeded that value a step-like graph is obtained as in *fig. 4*. The influence of the spread, as given expression in the step-like graph, is for the greater part eliminated when that line is replaced by a continuous curve and the required

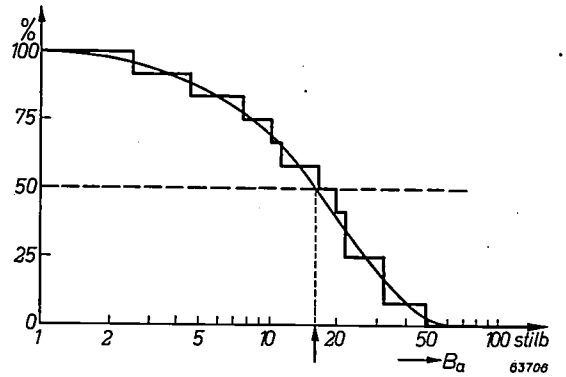


Fig. 4. For any lighting situation (ambient brightness B_s, solid angle ω under which the fitting is seen, viewing angle α) the adjustments for the permissible brightness B_a of the fitting as made by the various observers are arranged to form a step-like graph: the ordinate for each value of the brightness B_a indicates the percentage of observers judging that brightness to be permissible. The continuous curve smooths out the deviations among the observations seen in the step-like graph, the point of intersection with the 50% line giving the required "permissible brightness" for the lighting situation in question.

permissible brightness is read from the point where that curve intersects the 50% line. The more steps there are, that is to say, the larger the number of observers employed, the easier it is to draw the continuous curve. In our experiments all adjustments were made by 12 and in some cases even by 15 observers, a very much larger number than were employed for the investigations previously referred to. Of course, for each observer, in each measuring situation, again the average was taken of a number of adjustments made, starting from the side of low as well as high brightnesses.

Results of the tests

In view of the large number of parameters only an extract of the results can be given here.

Fig. 5 relates to the first series of experiments with a square source of light. Here the permissible brightness has been plotted as a function of the solid angle ω under which the lamp is seen, for three different values of the viewing angle α; the ambient brightness was 2×10^{-3} stilb (cd/cm^2), corresponding to the brightness of white paper under an illumination of 100 lux. The general trend of the lines answers the expectation: with increasing apparent size (ω) of the source of light the discomfort glare apparently becomes stronger and the permissible brightness accordingly less. This is in agreement, too, with the simple principle of the addition of glare

factors according to Harrison, since a large source of light may be regarded as being composed of two or more adjoining small sources of light.

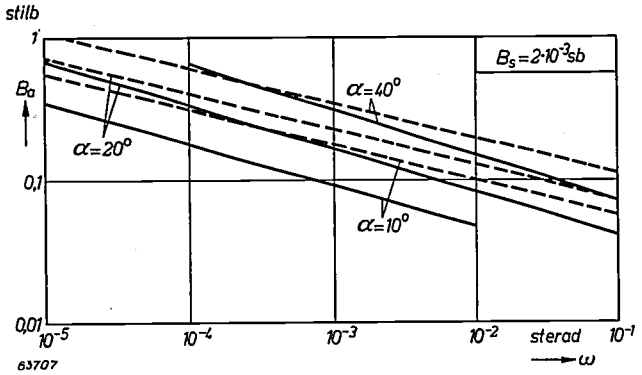


Fig. 5. Permissible brightness B_a of a square lamp fitting as function of the solid angle ω for three different viewing angles α. The corresponding lines according to Luckiesh and Guth [3]) are indicated in broken lines. Ambient brightness $B_s = 2 \times 10^{-3}$ stilb.

Also as regards the order of the values found there is reasonable agreement with former measurements, as appears from the dotted lines representing the results obtained by Luckiesh and Guth. This, however, is no longer true when we come to compare the results for small ambient brightnesses; see *fig. 6*, which applies for 0.04×10^{-3} stilb. According to our results, especially for large solid angles, the permissible brightnesses in this case are very much smaller than those according to Luckiesh and Guth, the difference being attributable mainly to the different criteria used. It seems to us, however, that all the same, judging from our measurements, the discomfort glare of the square (or round) light source, both with high and with low ambient brightnesses, can be described by a formula similar to (1), be it with slightly different powers (see the report quoted in footnote [6]).

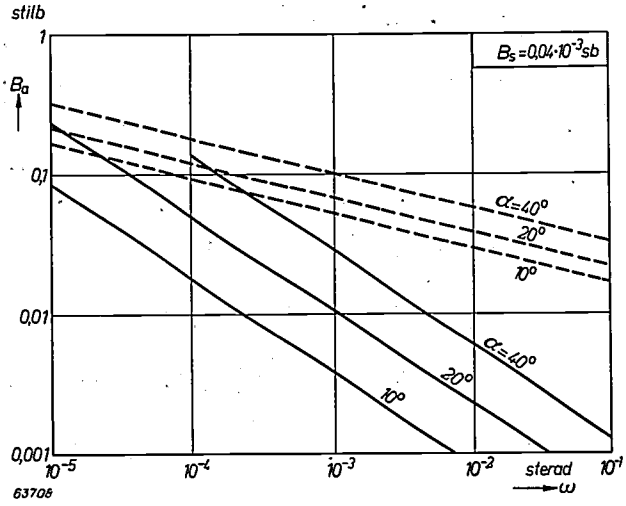


Fig. 6. The same as fig. 5 but for a much lower ambient brightness $B_s = 0.04 \times 10^{-3}$ stilb.

The results of the second series of tests, with elongated light source, showed more serious discrepancies. *Fig. 7* gives the permissible brightness again as a function of the solid angle ω and with the viewing angle α as parameter, with an ambient brightness of 2×10^{-3} stilb; the curves for a square source of light given in fig. 5 are reproduced here in fig. 7 in broken lines. The first thing to be noticed is that with an elongated source of light a very much higher brightness is permissible than in the case of a square or round one with the same visible surface! Thus we find confirmation of our surmise that it is not to be assumed that only the surface and not the shape governs the discomfort glare from a lamp. On the contrary, the shape appears to have a decided influence [7]). Physiologically the influence is to be so interpreted that the light striking the eye from the side makes it less sensitive to the glaring effect of the light in the central part of the field of vision.

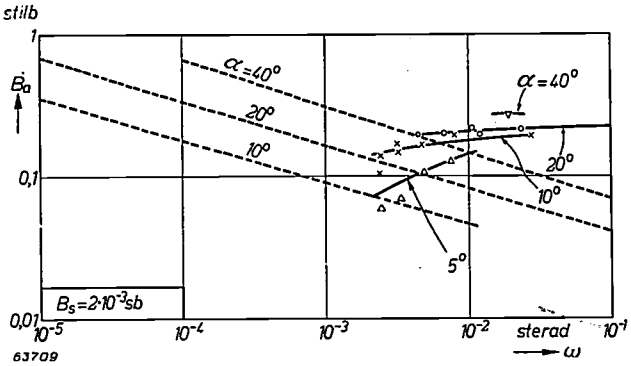


Fig. 7. Permissible brightness B_a of an elongated, horizontally suspended lamp fitting (visible surface 1.40 m × 0.10 m) as a function of the solid angle ω for different viewing angles α. The dotted lines are the corresponding lines from fig. 5 for a square fitting. Ambient brightness $B_s = 2 \times 10^{-3}$ stilb.

This effect appears to be so strong that in fact in fig. 7 (where the increased value of ω was obtained by bringing the observer nearer, thus with the light striking his eye more and more from the side) the permissible brightness increases with increasing solid angle ω. This entirely unexpected result goes to show that the simple method of adding up the glare factors of separate sources of light can certainly not be applied for lamp fittings that are not round in shape.

Finally we come to the series of experiments with naked fluorescent lamps and variable slit width.

---

[7]) From recent investigations conducted by P. Petherbridge and R. G. Hopkinson (Discomfort glare and the lighting of buildings, Trans. Ill. Eng. Soc. 15, 39-71, 1950, No. 2), where the shock method was likewise departed from in favour of a criterion somewhat analogous to ours, the shape of the light source proved to have a similar effect on glare.

Some of the results are represented in *fig. 8*, which has been constructed somewhat differently from the others: here the permissible slit width Δ has been plotted as a function of the distance d from the observer (cf. fig. 3), for a brightness of 0.4 stilb of the lamp ("TL" 40 watt) and an ambient brightness of 3×10^{-3} stilb; the lamp was suspended at a level 1.50 m higher than that of the observer's eye (from this datum the viewing angle α can be derived for any distance d). The general trend of the curve found again agrees with the expectation: the smaller the viewing angle α the more the observer has need of screening the light source (even if the reduction of α is accompanied by a reduction of ω). What strikes one as peculiar, however, is that practically the same curve is found for the three different lamp lengths of 0.60, 1.20 and 2.40 m (see the different measuring points). Thus we have here again a striking contradiction of the "glare factor" method, according to which with twice the length half the permissible width would have been expected.

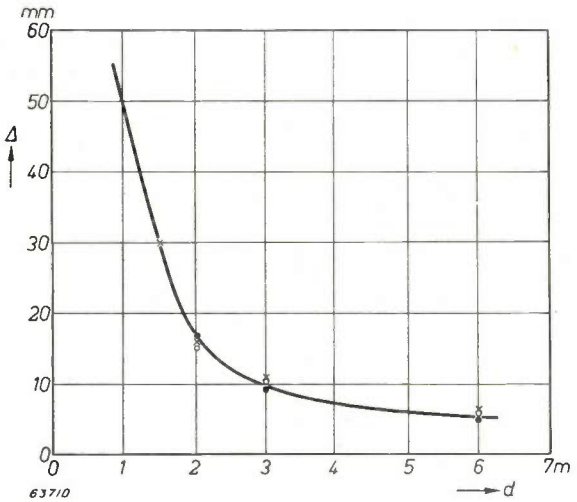


Fig. 8. Permissible visible width Δ of a fluorescent lamp in the set-up according to fig. 3, as a function of the distance d from the observer. The lamp is suspended 1.50 m above the eye level of the observer. Ambient brightness 3×10^{-3} stilb. The small circles refer to a lamp length of 0.60 m, the small crosses to one of 1.20 m and the dots to a length of 2.40 m.

Conclusion

It would be disappointing if from our experiments nothing else could be concluded than that a line of reasoning hitherto followed and in itself attractive may lead to erroneous predictions. The results reached have, however, indeed something positive about them. In the first place, even limited in scope as they were, these experiments can already serve for certain quantitave conclusions to be drawn as regards the most efficient construction of lamp fittings. It can be deduced, for instance, that the brightness of 0.3 stilb generally accepted as permissible for diffusers is really a factor 2-3 too high. When there are many diffusers, some of which at least the eye cannot escape, the lighting is certainly not comfortable. Further it may be deduced how the louvres have to be constructed in order that under any situations occurring (eye level and distance from the observer) the fluorescent lamps may not cause any feeling of discomfort. It is in accordance with these results that, among others, the louvres sketched in *fig. 9* have been designed, which are employed, for instance, in the fitting illustrated in *fig. 10*.

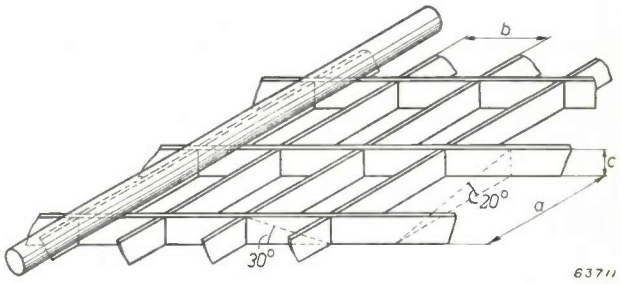


Fig. 9. Schematic diagram of a louvre. The "TL" lamps are mounted above the louvres in the direction shown (only one lamp has been drawn). The ratios of the dimensions a, b, c of the cells ensure complete screening of the naked lamps for viewing angles of respectively $\alpha < 30°$ and $\alpha < 20°$ (cf. fig. 2). Within reasonable limits one is free in the choice of the absolute dimensions. Also the pattern of the louvres can be altered on aesthetical or practical grounds provided the same screening angles are maintained.

Fig. 10. Top: Fitting for four "TL" lamps of 40 W provided with rhombic louvres with screening angles according to fig. 9. This type of fitting was employed on a large scale for the lighting of the new head offices of the Rotterdam Bank at Rotterdam.
Bottom: The same lamp fitting opened.

In the second place it is to be regarded as a positive result of our experiments that they do at least indicate the direction in which the reality deviates from the "glare factor" hypothesis. In fact, the conclusions that are to be drawn both from fig. 7 and from fig. 8 prove that the total discomfort glare from several light sources is less serious than would appear from an adding up of their separate glare effects. It has even been shown that in some cases the discomfort glare is reduced by the addition of further bright surfaces! The following actual fact taken from experience may be added to this statement. When an elongated source of light at right angles to the direction of vision and observed under a small viewing angle α has a brightness that is not disturbing, it appears that a number of identical sources of light with gradually increasing angle α can be placed in the field of vision simultaneously without causing the lighting to become uncomfortable. This is of great importance in practice, since it means that when illuminating a long hall with translucent "TL" fittings it is only necessary to make sure that the fittings farthest away from the observer do not cause any feeling of discomfort.

To obtain a full insight into the question of discomfort glare from several sources of light simultaneously present in the field of vision, more extensive experiments will, however, be needed with different configurations of the light sources.

Summary. The lighting installation in an office or suchlike space may be deficient not only on account of the level of illumination being too low but also by reason of an unsatisfactory distribution of brightness in the field of vision, as a consequence of which a feeling of discomfort may be caused, especially when one has to stay in the room for any length of time. From investigations into the suitable distributions of brightness ("brightness engineering") it appears, inter alia, that the fittings suspended in the room should not have too high a brightness. Previous investigations carried out in regard to the permissible brightness of fittings were based on Holladays's shock method, whereby the source of light was only momentarily shown to the observer and then his judgment of the glare, or of "discomfort" glare, was determined. In a series of experiments carried out by Philips this method has been departed from: the observers (there were at least 12, so as to get reliable results in spite of the rather great spread) were required so to adjust the brightness of a test lamp fitting themselves, under various conditions, that no feeling of discomfort would be felt when staying in the room a long time. The tests were performed with a square and an elongated source of light (the latter in imitation of a fitting with fluorescent lamps), whilst in a further series of experiments it was possible to investigate the dimensions required for the partitions in a fitting with louvres. Inter alia, these tests yielded the surprising result that the elongated light source may be allowed to have a greater brightness the greater the solid angle under which it is seen, such in contrast to the square source of light. From this it appears that Harrison's method, according to which the total glare effect of a number of sources of light is derived from the addition of their separate "glare factors" (equivalent surfaces), cannot hold for general application because it does not allow for the shape of the light sources. The results of these experiments permit of some conclusions being drawn that are of practical value.

RADIOGRAPHIC EXAMINATION OF ELECTRONIC VALVES

621.385:621.386.1:620.179

The demands made nowadays upon electronic valves for certain applications have led to the dimensions of the electrodes and their clearances being made smaller and smaller. Naturally the tolerances are then also very narrow and thus machining has to be most precise. For instance, in the case of a certain high-frequency pentode with coaxial cylindrical electrodes, an average eccentricity of 20 μ of the control grid with respect to the cathode is sufficient to change the valve characteristic to such an extent that the valve no longer meets the specification.

Formerly the cause of such a deviation from the prescribed characteristic was traced by removing the bulb of the valve and carefully detaching the anode and the surrounding screen with cutters and pliers, then inspecting and measuring the cathode and grid system. The latter was done with what is called a measuring projector, an optical apparatus with which a strongly magnified (up to 100 times) undistorted and bright silhouette of the object under examination is projected onto a frosted glass plate, the necessary measurements then being taken on this silhouette with a ruler.

This method of dismantling of the valve is not entirely satisfactory, as not only does it destroy the sample but there is always the risk of distortion of those parts which have to be measured. Nevertheless, under good circumstances, an accuracy of measurement of about 10 μ can be attained.

Recently a method has been developed which, while it gives the same accuracy of measurement as does the older method, avoids all possibility of distortion of the valve electrodes. By this new method it is possible to use X-rays as a means of projection.

To obtain good definition an X-ray tube with the smallest possible focus has to be used. When the object to be examined is placed at B (see *fig. 1*) at a distance v from the focus A, the dimension of which is Δ, then in a plane C at a distance b behind the object a shadow picture is formed with a geometrical blurring

$$\frac{b}{v}\Delta .$$

Details in the picture that are smaller than this blurring cannot be observed.

Since the projected shadow is $(b + v)/v$ times the size of the object, the smallest perceptible details of the object are given by:

$$\frac{v}{b+v}\Delta\,\frac{b}{v}=\frac{b}{b+v}\Delta .$$

In the set-up employed $v = 200$ cm,
$b = 4$ cm,
$\Delta = 0.3$ mm,

so that the resolving power amounts to 6 μ. The projected shadow is then only 2% larger than the object.

There are two ways of viewing the image: with the aid of a fluorescent screen or by means of a photographic recording. The sharpness of a fluorescent screen is quite inadequate, so that we are obliged to use the photographic method, although it takes more time.

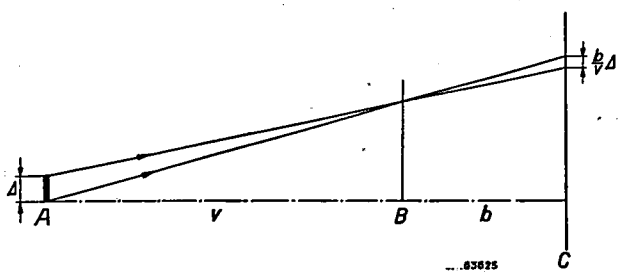
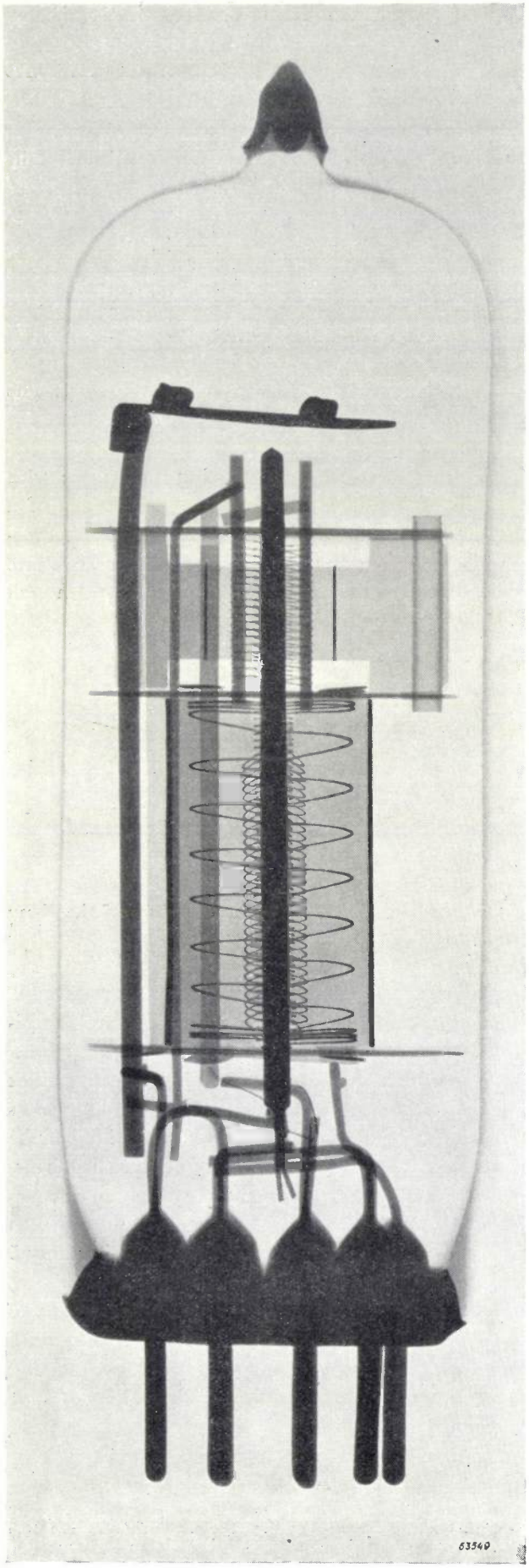


Fig. 1. Blurring of the X-ray image. A focus of the X-ray tube, B object plane, C image plane. With the set-up described here the object distance $v = 200$ cm, the picture distance $b = 4$ cm.

The normal photographic plate or film, emulsified on one side, with a grain of only a few microns, gives good results. An intensifying screen cannot be used because this would reduce the sharpness too much. A special fine-grained material can be used and gives a greater accuracy of measurement, but since it takes a rather long time to develop (15 to 20 min) a normal plate is mostly used so as to save time, the developing time then being only 5 minutes.

With the set-up now in use a "Rotalix" tube of the type O 75 is employed, this being a tube with two foci, but for our purposes only the smaller focus is used (apparent focus dimensions 0.3 × 0.3 mm [1]).

[1]) For a description of this tube see G. C. E. Burger, B. Combée, and J. H. van der Tuuk, X-ray fluoroscopy with enlarged image, Philips Techn. Review 8, 321-329, 1946.



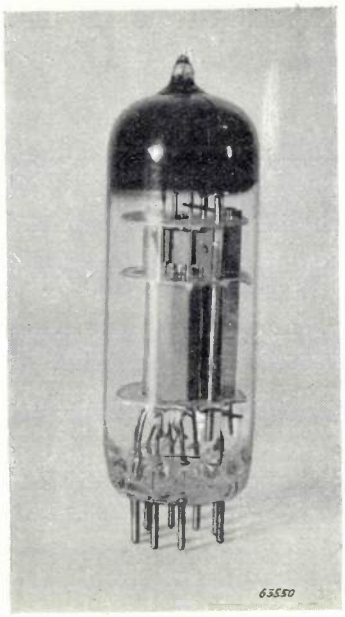
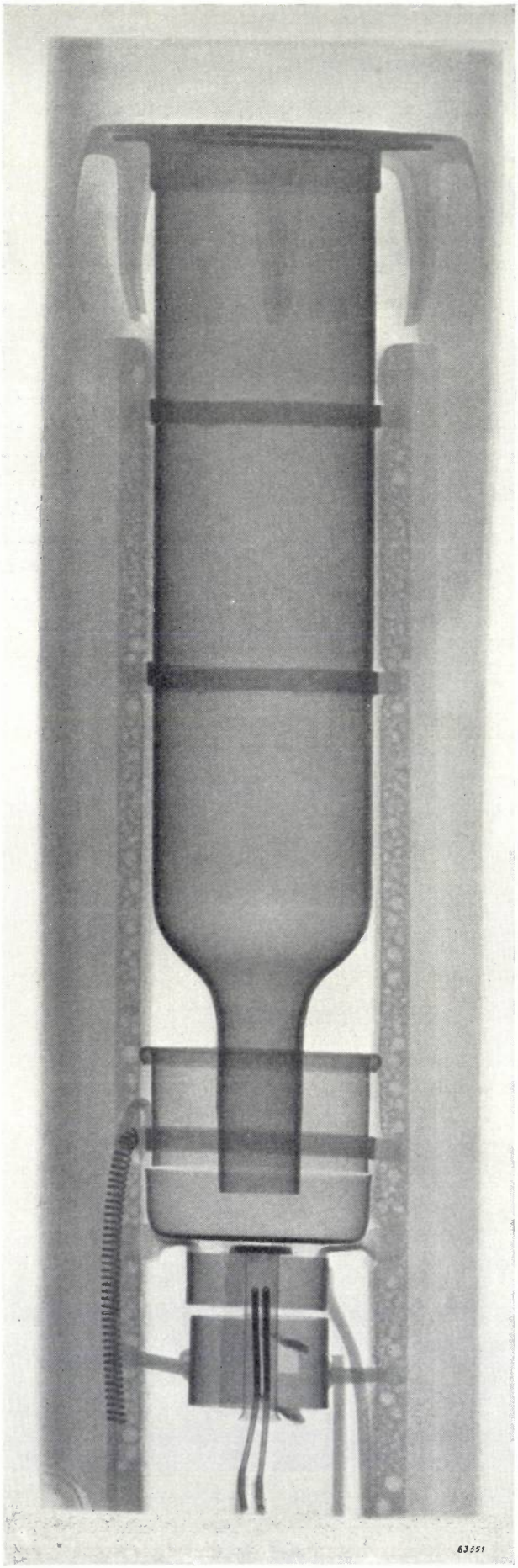


Fig. 2 Combined pentode and triode for television reception, type ECL 80.
Right: Natural-size photograph. *Left:* X-ray photograph enlarged 4 times. The three grids of the pentode are clearly seen in the lower part of the valve. The cylindrical anode is distinguished by the darker field behind the grids and the two lines running down the side of them. The indirectly-heated cathode is concealed by the supporting wires of the grid, but the coiled filament can be seen at the bottom of the valve. The second grid, which does not reach right to the top, has apparently been distorted at the top end when the winding jig was removed and this was the cause of the characteristic deviating beyond the tolerance and the valve being rejected on the testing bench.

The X-ray tube, the object and the film are screened off with an iron case lined with lead. On the front is a fluorescent screen on which a silhouette of the valve being examined can be seen, so as to be able to set it in the right position. The valve is positioned by means of three knobs outside the case, with which the valve can be turned in two directions at right angles to the X-ray beam and also moved vertically. The mechanical transmission consists of three flexible spindles. The operating panel for the X-ray tube is situated underneath the fluorescent screen and is fitted with a time switch.

For studying the behaviour of the valves while heating up provision has been made for connecting them to the working voltages.

The exposure time varies from 1 to 2.5 minutes according to whether the valves being examined have lime-glass or lead-glass bulbs. Since it is not usually necessary to keep the photograph, the fixing and rinsing is done quickly and the photograph then rapidly dried with heated dry air. In this way it is possible to finish taking the measurements 30 minutes after starting operations, a factor of



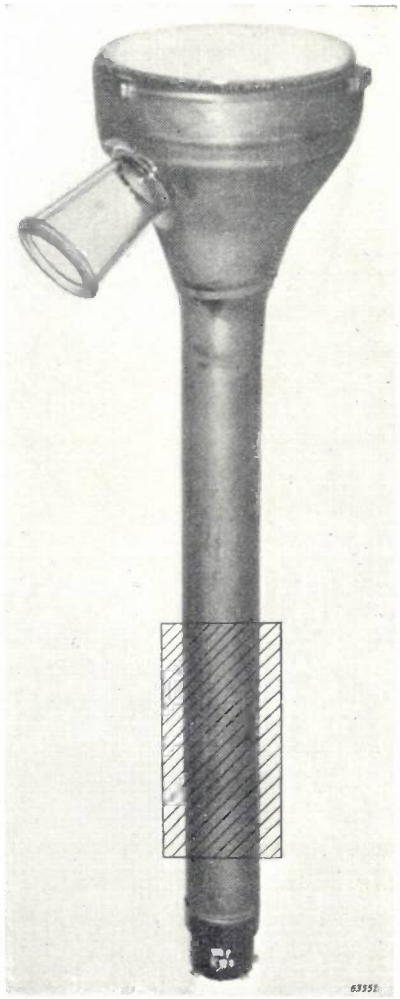


Fig. 3. Cathode-ray tube for television projection, type MW 6-2. Right: photograph of the tube (half the actual size). Left: X-ray photograph of the electron gun (hatched part in the photograph on the right) enlarged 3.5 times.

great importance when it is necessary to intervene quickly in the production process.

The photographs are measured up under an optical projector in the same way as was formerly done with the dismantled systems of the valves.

The accuracy of the measurements can very easily be checked by photographing a gauge together with the objects. Practice has proved that a total accuracy of 10 μ is reached.

The photographs reproduced here give an idea of the sharpness obtained, though naturally much of it is lost in the reproduction.

H. B. van Wijlen

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Nederlands.

1921: J. D. Fast: Ageing of iron and steel (Iron and Coal Trades Review **160**, 837-844, April 1950).

A study has been made of parts played by oxygen, nitrogen and carbon in the ageing of unalloyed iron and in the ageing of iron containing manganese. For this purpose, pure iron was first prepared and known quantities of impurities were added, separately or together.

Quench-ageing: As shown by Vickers hardness determinations, carbon and nitrogen cause significant quench-ageing in iron that is otherwise pure. On the other hand, the quench-ageing of oxygen-containing iron is very slight. The addition of 0.5 per cent Mn has no effect on the action of carbon but practically suppresses the quench-ageing of nitrogen-containing iron. This remarkable effect of manganese has been confirmed by determinations of damping. In the case of the damping of torsional oscillations set up in wires, the maximum due to nitrogen is broadened and displaced towards higher temperatures by the addition of 0.5 per cent Mn. Thə experiments indicate that the nitrogen atoms are situated preferentially in the interstitial positions which are in the immediate neighbourhood of the manganese atoms. The observed damping shows that the nitrogen atoms can make jumps around the atoms of manganese, but that they find it difficult to leave these atoms. Heating for some tens of hours at 200 °C is necessary to precipitate a noticeable quantity of nitrogen in the alloys containing manganese.

Strain-ageing: Oxygen does not cause strain-ageing in iron. The essential cause of this phenomen is nitrogen, less than 0.001 per cent of which suffices to produce the maximum strain-ageing. Even at room temperature this ageing takes place with great speed. Strain-ageing due to carbon only appears at a noticeable speed at elevated temperatures, such as 100 °C. This difference between the effects of carbon and nitrogen apparently depends principally on the difference in their solubilities in iron. The time required for strain-ageing is controlled mainly by the product of the diffusion coefficient and the solubility of carbon or nitrogen in iron. Manganese has no appreciable effect on the strain-ageing caused by carbon or nitrogen.

1922: N. W. H. Addink: A rapid and accurate method of measuring line intensities in spectrochemical analysis (Spectrochimica Acta 4, 36-42, 1950, No. 1).

A method of measuring line intensities in spectrochemical analysis is developed, based on the principle of visual comparison with standard density lines (standard paper density scale). The method gives results as accurate as those obtained by means of the densitometer (5%): it includes a correction for background intensity and it speeds up the quantitative analysis, as the qualitative inspection of the spectrum and the measurement of line intensities can be made simultaneously. The use of a densitometer is eliminated. Examples of analysis results and of the time necessary for quantitative analysis according to this method are given.

1923: C. J. Dippel: The metal-diazonium process (Photographic Journal **90 B**, 34-41, 1950, No. 2).

A general survey is given of the principal characteristics of the new metal-diazonium reproduction process: the transformation of a latent image composed of an organic light decomposition product into a latent image composed of mercury droplets, followed by an intensification by means of physical development into a silver image. The characteristic properties of this new process are described and its high resolving power of 1,200 lines per mm is stressed.

1924: J. M. Stevels: Quelques nouveautés dans les recherches sur le verre (Verres et Réfractaires 4, 3-9, Febr. 1950, No. 1). (Some novelties in glass research.)

Description of sintered glass and its applications (see Philips Techn. Review **8**, 2-7, 1946) and of scale-glass (Philips Res. Rep. **1**, 129-134, 1946). The latter consists of small scales having a surface area af obout 1 mm^2 and a thickness of 1 to 5 μ, which are superimposed by sedimentation in a suitable liquid and which hold together by cohesion forces. It can be deformed while wet, whereas in a dry state it may replace mica. Finally a number of low-melting glasses are mentioned, having a high resistivity and low dielectric losses, e.g. a

glass consisting of 45 SiO_2, 32 PbO, 5 CaF_2, 4 Na_2O and 14 K_2O (percent by weight), which can be sealed to iron. Some properties (resistivity and loss angle) as a function of temperature and frequency are shown in the form of graphs.

1925: G. Diemer and J. L. H. Jonker: Secondary-emission valve as wide-band amplifier for decimeter waves (Wireless Eng. **27**, 137-143, May 1950).

An experimental secondary-emission valve is described. The high figure of merit ($g_m/C = 3.0$ mA/V-pF) that is obtained by adding one stage of secondary emission to a grounded-grid triode of rather conventional construction makes the valve useful as a wide-band amplifier for those cases where there is no need for a very low noise figure; typical figures are: at 1m wavelength 30 db gain (G) with a bandwidth (B) of 3.5 Mc/s, at 50 cm $G =$ 15 db with $B = 20$ Mc/s, at 30 cm $G = 10$ db with $B = 10$ Mc/s. The maximum power output is for $\lambda > 1$ m about 1.5 watts. At 7 m wavelength the noise figure amounts to 12 db; this rather high value is due to the secondary emission. It is shown that for this secondary-emission noise a kind of space-charge smoothing effect exists.

1926: E. W. Gorter: Saturation magnetization of ferrites with spinel structure (Nature, London **165**, 798-800, May 20, 1950).

The magnetic moment per molecule M_s of components $x\ ZnFe_2O_4 + (1 - x)\ Me^{II}Fe_2O_4$, where Me^{II} is a divalent metal, are measured at liquid-oxygen and liquid-nitrogen temperatures and plotted as a function of x. The curves show a maximum in the neighbourhood of $x = 0.4$, the tangents at $x = 0$ converge to the value $M_s = 10$ Bohr magnetons at $x = 1$. The results are in agreement with Néel's theory.

1927: H. G. Beljers and D. Polder: g-factors in ferrite materials (Nature, London **165**, 800, May 20, 1950).

The g-factors of a number of ferrites (see No. **1926** of these abstracts) have been determined by studying the absorption of microwaves in small spherical samples of the material (see No. **1850** of these abstracts).

1928: J. L. Snoek: Grain-boundary slip and magnetic relaxation at high temperatures in iron (Physica, The Hague **16**, 336, 1950, No. 3).

Measurements of Fahlenbrach on magnetic relaxation at high temperatures (400 °C to 600° C) in iron and silicon iron are compared with measurements of the internal damping (400 °C to 600 °C) of fine-grained samples of iron (grain size 3μ) by Kê. Both phenomena are ascribed to grain-boundary slip. This gives additional support to the author's theory of the room-temperature after-effect, which makes elastic after-effect in the Bloch boundary zone solely responsible for the observed time lags.

1929: J. L. Snoek: The weak ferromagnetism believed to be present in α-Fe_2O_3 and other antiferromagnetic compounds (Physica, The Hague **16**, 333-335, 1950, No. 3).

The evidence for a small permanent moment observed below the critical temperature in most antiferromagnetic compounds is traced back to lattice imperfections.

1930: F. A. Kröger and J. Dikhoff: Trivalent cations in fluorescent zinc sulphide (Physica, The Hague **16**, 297-317, 1950, No. 3).

Incorporation of monovalent cations in a lattice consisting of divalent ions is only possible to an appreciable extent when the lack of positive charge resulting from the substitution of a monovalent cation for a divalent one is compensated. This compensation can be achieved by a simultaneous incorporation of monovalent anions, or of cations of a valency higher than two. On this basis it is explained that ZnS becomes fluorescent by the monovalent activators Ag^+, Cu^+, Au^+ and Zn^+ when halogens or trivalent cations are present. Some of the trivalent ions incorporated in this way are found to cause effects of their own (electron traps, fluorescence, killing of fluorescence due to the other centres). An atomic model of the centres of fluorescence is given.

1931: F. A. Kröger and N. W. Smit: The physical chemistry of the formation of fluorescence centres in ZnS-Cu (Physica, The Hague **16**, 317-328, 1950, No. 3).

On the basis of a model for the centres of fluorescence in ZnS-Cu arrived at in previous publications (Nos. **1902** and **1930**), the formation of centres with products prepared in controlled atmospheres of H_2S-HCl is discussed, assuming equilibrium to exist between the solid and the gas phase.

R 130: B. D. H. Tellegen and E. Klauss: The parameters of a passive four-pole that may violate the reciprocity relation (Philips Res. Rep. **5**, 81-86, 1950, No. 2).

Properties at a fixed frequency of linear, passive four-poles that may violate the reciprocity relation

are investigated. Necessary and sufficient conditions are derived for the four-pole parameters in order that the four-pole may be passive. For a four-pole containing only one resistor or one only reactor a certain connection exists between these parameters.

R 131: C. J. Bouwkamp: On the characteristic values of spheroidal wave functions (Philips Res. Rep. **5**, 87-90, 1950, No. 2).

The first few terms of a power series expansion are given for the characteristic values of spheroidal wave functions of integral order and integral degree. Some numerical results are communicated for functions of order one.

R 132: W. F. Dil: Principles of measurements on coupled circuits (Philips Res. Rep. **5**, 91-115, 1950, No. 2).

An analysis is given of an electrical circuit consisting of a valve and two coupled circuits, the coupling of which is capacitive as well as inductive and resistive. The behaviour of this circuit is characterized by the resonance curve and the gain at the central frequency. The conditions for symmetrical resonance curves are investigated. A survey is given of the most commonly employed methods for measuring the characteristic quantities such as coupling factors and qualities of the circuits.

Finally, new measuring methods are discussed which are practically free of the well-known difficulty of disturbing the normal operating conditions of the circuits, thereby affecting the accuracy of the measurement. Moreover, these methods offer the possiblity of measuring all the characteristic quantities of a system of coupled circuits having a symmetrical resonance curve.

R 133: R. Dorrestein: On the energy-flow distribution in certain types of paraxial beams (Philips Res. Rep. **5**, 116-127, 1950, No. 2).

In a paraxial beam proceeding along the axis of a rotationally symmetric (light-optical or electrostatic) refractive medium, special pairs of cross sections may exist, in which the corresponding distributions of energy current density (or intensity) are mutually independent. They are called "independent" cross sections. From the intensity distributions in two such independent cross sections, one can calculate the intensity distributions in any other cross section by means of a suitable composition-product integral. The "Gaussian beam" is defined by a Gaussian intensity distribution and a Gaussian angular energy-flow distribution in one cross section. This class of beams possesses an infinite number of pairs of independent cross sections. The equation for the effective radius r, as a function of the distance z along the axis, is identical with the paraxial-ray equation for $r(z)$ in cylindrical coordinates z, r, φ.

R 134: R. Dorrestein: Note on the image formation in cathode-ray tubes (Philips Res. Rep. **5**, 128-130, 1950, No. 2).

The theory of Gaussian beams (see No. **R 133** of these abstracts) is applied to the electron beam in a cathode-ray tube with a convential type of gun. It appears that in principle neither the paraxial crossover, nor the cathode image, but rather some intermediate point is to be imaged on the screen in order to yield the smallest possible spot. However, in practical cases this intermediate point nearly coincides with the crossover, which is in agreement with the generally accepted concept of image formation in cathode-ray tubes.

R 135: K. S. Knol and G. Diemer: Theory and experiments on electrical fluctuations and damping of double-cathode valves (Philips Res. Rep. **5**, 131-152, 1950, No. 2).

A calculation is given of the internal resistance and the noise of symmetrical double-cathode valves containing two hot cathodes opposite each other. The results of measurements with indirectly heated oxide-coated cathodes agree with these calculations and show that, in contradiction to a theory of Fürth, the equivalent noise temperature of such valves does not exceed the true cathode temperature. Directly heated valves show small deviations from the theory, which are discussed.

R 136: G. Diemer and K. S. Knol: Low-level triode amplifier for microwaves (Philips Res. Rep. **5**, 153-154, 1950, No. 2).

It is shown that — with a new type of cathode (the so called L-cathode, see Philips Techn. Review **11**, 341-350, 1950, No. 12) — a low-noise triode for centimetre-wave amplification can be constructed, giving rise to an overall noise figure of 7 db.

VOL. 12 No. 8, pp. 213-240 FEBRUARY 1951

Philips Technical Review

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

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

THE OXIDATION OF HEAT-RESISTANT ALLOYS IN THE PRESENCE OF FOREIGN OXIDES

by J. L. MEIJERING and G. W. RATHENAU.

669.018.44:620.196:546.776-31

The construction of modern thermal engines with their high working temperatures has been made possible by the development of heat-resistant alloys possessing the mechanical properties required. Nevertheless cases still cocur where parts of these constructions show an unexpectedly rapid corrosion. From a detailed study of such phenomena of corrosion the authors have succeeded in throwing some light upon their causes.

In recent years more and more interest has been shown in heat-resistant alloys. These alloys have particularly become important for the construction of modern thermal engines (gas turbines, jet engines, hot air engines), in view of the fact that in order to reach a high efficiency the metallic parts of these engines must withstand working temperatures above 700 °C. This means that the materials to be used have to satisfy the requirement that, at the chosen working temperature and under the chosen mechanical load, during a reasonable lifetime for the engine particularly the plastic deformation and the corrosion due to chemical reaction with the surroundings do not exceed the given tolerances. This article will deal with a certain aspect of these phenomena, and in particular with the corrosion of heat-resistant metals and alloys through oxidation.

In principle there are two ways of avoiding oxidation of metallic parts of constructions at elevated temperatures.

In the first place "precious" metals can be chosen which, with respect to their surroundings, have too little affinity to oxidize. This means that at the working temperature the oxygen pressure (dissociation pressure) of the oxides that might be formed from the metal is greater than the partial pressure of oxygen in the atmosphere surrounding the metal. For instance, the oxygen pressure of silver oxide Ag_2O above 150 °C is more than $^1/_5$ atmosphere, so that above that temperature silver does not oxidize in air of atmospheric pressure.

However, owing to their unfavourable mechanical properties and high prices, precious metals (or alloys thereof) are hardly ever used for heat-resistant constructional parts. Most lythe second course is followed, basic or alloy elements being chosen which have the property of forming on the metal dense and strongly adhering oxide skins practically impervious to oxygen- and metal-ions. The alloy underneath such a skin certainly has the tendency to oxidize at the working temperature, but the rate at which this oxidation takes place is reduced to sufficiently small proportions by the obstruction to diffusion. We could express it in this way, that in the case of precious metals it is the thermodynamics (position of the dissociation equilibrium of the oxide) that make the material resistant to oxidation, whereas in the case of the protective oxide skin only the kinetics of the oxidation are affected in a sense favourable for the purpose.

As regards the protection afforded by an oxide skin, in many cases occurring in practice account has to be taken of the fact that at the surface the material comes into contact not only with oxygen or air but also with foreign substances, for instance oxides having an entirely different composition than that of the oxide skin making the material resistant to oxidation. It has long been known that such a contact sometimes considerably reduces this resistance to oxidation. An example of this is given in electric heating wires surrounded by certain ceramic materials. The smallest trace of asbestos in

such a material may cause a wire of FeCrAlCo 73/20/6/1 [1]) ("Kanthal") heated to 1300 °C to fuse in a few seconds, whereas the normal lifetime of this wire at that temperature is some hundreds of hours. Another example is the abnormal corrosion of stainless steel in combustion chambers fired with fuel oil [2]). In this case it is the vanadium oxide so often present in such oils that is the cause of the trouble; this oxide may amount to 80% of what is left from the combustion of the oil. A third example relates to stainless steels containing molybdenum, from which, at the working temperature, volatile molybdenum oxide MoO_3 is formed [3]). Here again there is abnormal corrosion, which is attributable to the reaction of the oxide skin of the metal with the MoO_3 vapour.

Similar phenomena have been encountered by the authors when investigating in the Philips Laboratories at Eindhoven the oxidation-resisting properties of heat-resistant copper alloys and steels at 750 °C in air strongly contaminated with MoO_3. In the case of aluminium bronze (AlCu 8/92), for instance, heated in a furnace contaminated with MoO_3 vapour, after brushing off the oxide formed, which adhered loosely to the surface, a loss in weight was measured which was many hundreds of times greater than the loss found after oxidation in pure air. The graph in *fig. 1* shows that after three hours in the contaminated atmosphere the loss in weight already amounted to 6 mg/cm², as compared with about 0.01 mg/cm² in pure air. A remarkable fact shown by this graph is that the corrosion continues at a rapid rate when the material is further heated in a pure atmosphere (thus not containing MoO_3). We shall revert to this detail later (see fig. 5*b*).

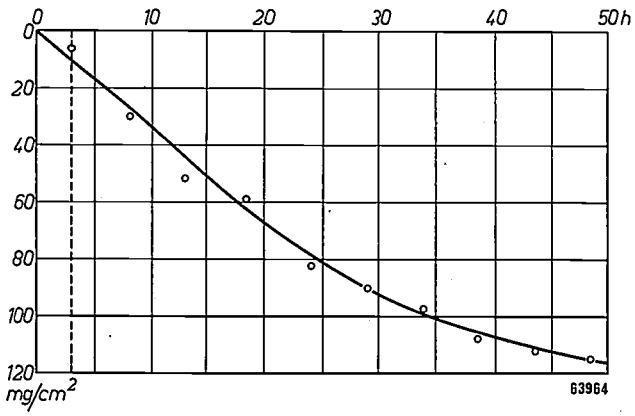


Fig. 1. Oxidation of a strip of aluminium bronze, AlCu 8/92 (cf. footnote [1])), heated at 750 °C in air contaminated with vapour from molybdenum oxide MoO_3. The loss of weight, measured after brushing off the loose oxide, is plotted as a function of time. — After three hours' heating the strip was transferred into an atmosphere of pure air; the oxidation appeared to continue at about the same rate.

With a view to gaining a general insight into the manner in which the oxidation of metals is influenced by contact of the protective oxide skin with foreign substances, in particular oxides, a series of systematic experiments have been carried out in this laboratory. In all cases MoO_3 was used as foreign oxide, but, as will appear from what follows, the results and the conclusions to be drawn therefrom may be regarded as more or less representative for the mechanism of corrosion also when other foreign oxides are active, and probably also in the case where a metal is corroded not by air but by other gases, as for instance chlorine.

Quantitative determination of the degree of oxidation

First of all the process of oxidation was investigated, by different means, as a function of temperature. It very soon became evident that abnormal oxidation of heat-resistant alloys caused by MoO_3 may occur already at relatively low temperatures, even below 500 °C. This was established by bringing samples of the metals and alloys into intimate contact with MoO_3 powder during the heating. Contact only with air containing MoO_3 does not, at these temperatures, cause any appreciable rapid corrosion, since the vapour pressure of MoO_3 at 500 °C is only about 10^{-5} mm Hg. For a proper mutual comparison, in all the tests the samples were therefore packed in MoO_3 powder and then heated in air. Reference samples of the same composition, carefully separated from the others, were heated to the same temperature in pure air alone.

In order to follow quantitatively the conversion through oxidation, in the test to which fig. 1 refers the loss in weight of the metal was taken as the measure. It appears, however, that in some cases this leads to difficulties, because sometimes it is difficult to remove the oxide skin entirely and, furthermore, the abnormal corrosion is often only local (as will be understood from what follows). Both these factors cause the results to be spread over a wide range, as is clearly seen from *fig. 2*, where the loss in weight measured with a series of strips of the same aluminium bronze as used for the test of fig. 1 — each strip heated for 3 hours in contact with MoO_3 and air — has been plotted as a func-

[1]) An alloy with 73 wt % Fe, 20 wt % Cr, 6 wt % Al, 1 wt % Co. Further in this article the abbreviated form will be used in every case.

[2]) P. Schläpfer, P. Amgwerd and H. Preis, Schweizer Archiv 15, 291, 1949.

[3]) W. C. Leslie and M. G. Fontana, Trans. Am. Soc. Metals 41, 1213, 1949.

tion of the temperature. It is seen that round about 480 °C oxidation strats to take place much more quickly, but from results with such a considerable spread it is difficult to draw more detailed conclusions as to the relation between the rate of oxidation and temperature.

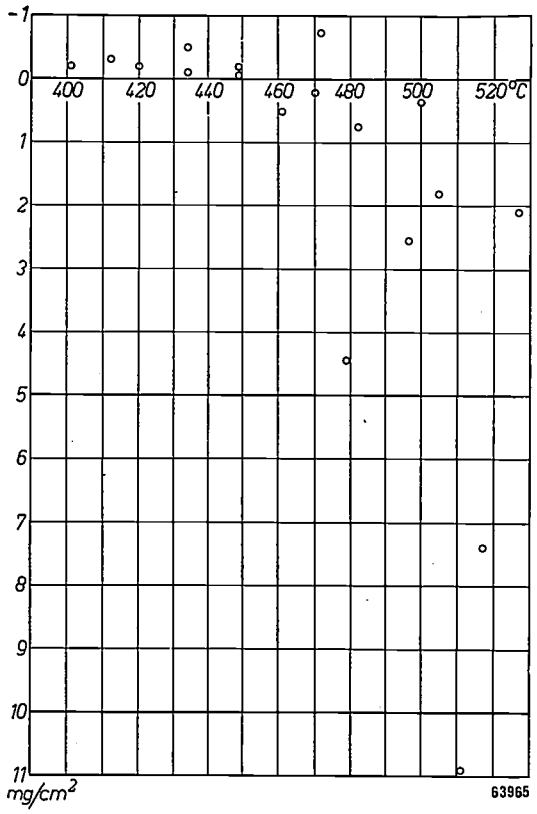


Fig. 2. Loss in weight of strips of AlCu 8/92 packed in MoO_3 powder and heated for three hours in air at different temperatures.

A more reliable and more practicable method for determining the degree of oxidation was found to be the following.

From each material to be investigated a set of five wires were used (diameter $2r_0 = \frac{1}{2}$ mm) and brought to a reproducible, stress-free, initial state by heating in a pure mixture of nitrogen and hydrogen. After oxidation in contact with MoO_3 and air, the average breaking strength (K_M) of the wires was determined at room temperature with a normal tensile testing machine. For comparison an identical set of wires were heated in pure air for the same length of time and at the same temperature, after which their average breaking strength (K_A) was likewise measured. As a measure for the abnormal oxidation we have now defined an "effective penetration depth" d:

$$d = (1 - \sqrt{K_M/K_A})\, r_0 . \quad . \quad . \quad . \quad . \quad (1)$$

The definition is based upon the assumption that the wires subjected to the tensile test always break at the same tensile stress and that no appreciable notch effects occur. This means to say that all wires break at one and the same value of the quotient of breaking strength and area of the remaining cross section of the wire. If the oxidation round the circumference of the wire penetrates to a depth d then the radius r_0 of the effective wire cross section is reduced to $r_0 - d$, from which the above formula follows directly (provided, as is indeed found to be the case, corrosion in pure air is negligible).

The penetration depth d relates to the most corroded spots, since a test specimen usually breaks at the thinnest place. Therefore, although in some places the oxidation may be impeded through incidental causes (see below), with this method of measuring and this definition reproducible results may still be expected. Incidentally, it may be noted that the penetration depth measured by the breaking strength is quite a good measure for corrosion in practice, since in practical engineering it is just the deterioration in mechanical strength of constructional parts, be it only local, that counts most. For a good understanding of the mechanism of corrosion, however, it is certainly also of importance to know whether and in how far there is question either of all-round corrosion or of a local, e.g. inter-crystalline, penetration. For this reason also cross sections of the wires were examined microscopically.

The fact that really reproducible values are found for the penetration depth is demonstrated by *fig. 3*, showing the penetration depth d determined according to eq. (1) after 3 hours' heating and plotted as a function of temperature, for the same aluminium bronze AlCu 8/92 to which figures 1 and 2 relate. Here the measured points show but little spread and the curve determined by the points clearly shows two discontinuities, the first at 470 °C and the second round about 535 °C.

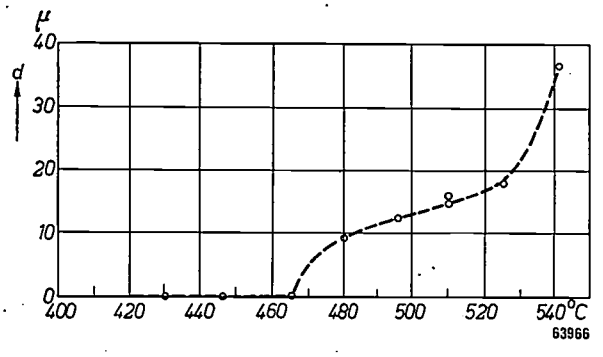


Fig. 3. Penetration depth d found from tensile tests with wires of AlCu 8/92 oxidized (in the same way as the strips of fig. 2) by heating for three hours in MoO_3 powder and air, at different temperatures.

Liquefaction at and in the protective oxide skin

The fact that at very definite temperatures the rate of oxidation changes discontinuously gives rise to the thought that phase transitions, namely melting or a change in the crystal structure, will play a part. In itself this is no new idea. The possibility of liquefaction playing a part in accelerated oxidation has been suggested before [4]). The examples given in the beginning also point in that direction: asbestos, which may cause rapid corrosion of heating wires, melts at about the temperature of 1300 °C mentioned; vanadium oxide, which occurs in fuel oil, melts already at about 650 °C. The melting point of molybdenum oxide lies, it is true, much higher than the temperature of about 500 °C at which abnormal corrosion was found, namely at 795 °C; but even so liquefaction is possible at those lower temperatures if low-melting eutectics occur.

To put this idea to the test we have determined the eutectic temperatures of the oxide systems concerned, with the aid of temperature-time curves (when a melt is cooled a retardation in the decrease of the temperature points to a separation of crystals, and if the cooling comes to a standstill for a time this points to a eutectic temperature having been reached, at which the whole of the liquid solidifies, as can in fact be seen to take place). In the case of the aforementioned aluminium bronze in combination with MoO_3 we found that the binary system [5]) MoO_3-Cu_2O has a eutectic temperature of 535 °C. The eutectic temperature of the ternary system MoO_3-MoO_2-Cu_2O proved to be 470 °C (that of MoO_3-MoO_2 lies at 778 °C).

Not only do we find here, therefore, very low melting eutectics, but the striking agreement between the eutectic temperatures — thus the lowest temperatures at which liquids can occur at the surface of the alloy — and the discontinuities in fig. 3 clearly lends support to the hypothesis that a causal relation exists between the formation of liquid and the markedly accelerated corrosion [6])[7]).

How is the accelerating effect of the formation of liquid upon the process of oxidation to be explained? It has to be imagined that the liquid, even if present in only small quantities, forms continuous channels along the grain boundaries of the protective oxide, and that owing to diffusion and convection in these liquid channels, which permit a much more rapid exchange of metal atoms and especially also of oxygen atoms than is possible in a solid phase, the protection afforded by the oxide skin becomes illusory. It is quite feasible that such channels may be formed. The interfacial energy of two adjacent grains of a crystalline phase (the oxide skin) is usually so great that the presence at the grain boundaries of a liquid phase which easily adapts itself to the local atomic structure leads to a reduction of the total interfacial energy. A striking example of this is the well-known case of copper immersed in molten bismuth, the liquid penetrating along the surfaces of the adjacent grains deep into the metal and thereby making it extremely brittle [8]). In many other cases, where the interfacial energies at the grain boundaries may perhaps be less, the liquid may anyhow penetrate along the ribs where three grains come together.

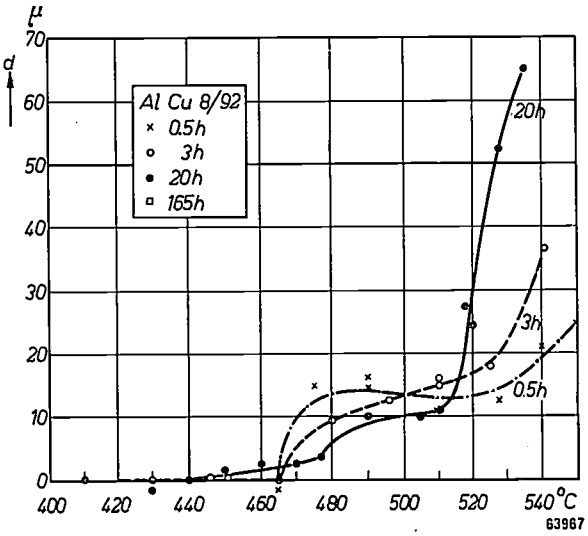


Fig. 4. The same as in fig. 3 but also after ½ hour, 20 hours and 165 hours' heating respectively. At not too high temperatures there is a peculiar dependency of the penetration depth upon the heating time. This is accounted for by a notch effect at the beginning of the oxidation.

Details of the formation of the eutectics

In the case dealt with in the foregoing, it is peculiar that the low-melting eutectic is not formed by interaction of the added oxide MoO_3 with the "protecting" oxide Al_2O_3 but by reaction with oxides of the basic metal, copper. (The oxidation of pure aluminium is not accelerated by MoO_3!) The possibility of such a reaction is probably to be ascribed to the fact that in the oxide skin of the AlCu oxidized in air there

[4]) See, e.g., W. Hessenbruch, Metalle und Legierungen für hohe Temperaturen, I., J. Springer, Berlin 1940. Strangely enough this idea has receded into the background again, as is evident from the article quoted in footnote [3]).

[5]) The influence of Al_2O_3 upon the eutectic temperatures is negligible.

[6]) J. L. Meijering and G. W. Rathenau, Nature, London 165, 240, 11th Febr. 1950.

[7]) G. W. Rathenau and J. L. Meijering, Metallurgia, 42, 167-172, 1950. This article gives a more detailed report on the investigations dealt with here, in particular also on tests with various alloys not mentioned here and on the results of the metallographic examination.

[8]) C. S. Smith, Trans. A. I. M. E. 175, 15, 1948.

are also Cu-oxides present, whereas the MoO_3 could also react directly with the Cu from the alloy underneath through cracks in the protective layer of oxide. If such capillary cracks do indeed play a part then it is understandable why the reaction is often localized, as mentioned above in connection with the definition of the penetration depth [9]). This may also be the explanation of the fact that, curiously enough, the penetration depth at first appears to decrease with increasing duration of heating. *Fig. 4* illustrates this effect. If, owing to its starting from fine cracks in the oxide skin, the corrosion at the beginning is greatly limited locally, then it lowers the breaking strength disproportionately owing to notch effect. After a greater length of time the corrosion becomes more uniform, as *figs 5a* and *b* show.

The influence of MoO_3 upon the oxidation of pure Cu is to be deduced from *fig. 6*. In this case it is of course one of the protecting oxides itself that reacts with the foreign oxide, namely Cu_2O (the eutectic temperature for CuO with MoO_3 is 735 °C). The influence of the eutectic point at 535 °C of MoO_3-Cu_2O is quite clear. Here the ternary eutectic point at 470 °C of MoO_3-MoO_2-Cu_2O is not manifest, though its influence is to be seen in measurements of the loss of weight, which are more sensitive than those of the penetration depth, when the object is to detect the very beginning of accelerated oxidation. The fact that here the formation of MoO_2 from MoO_3 is not noticeable, whereas it is so in the case of Al-Cu alloys, is to be accounted for by the strongly reducing action of Al (great stability of Al_2O_3).

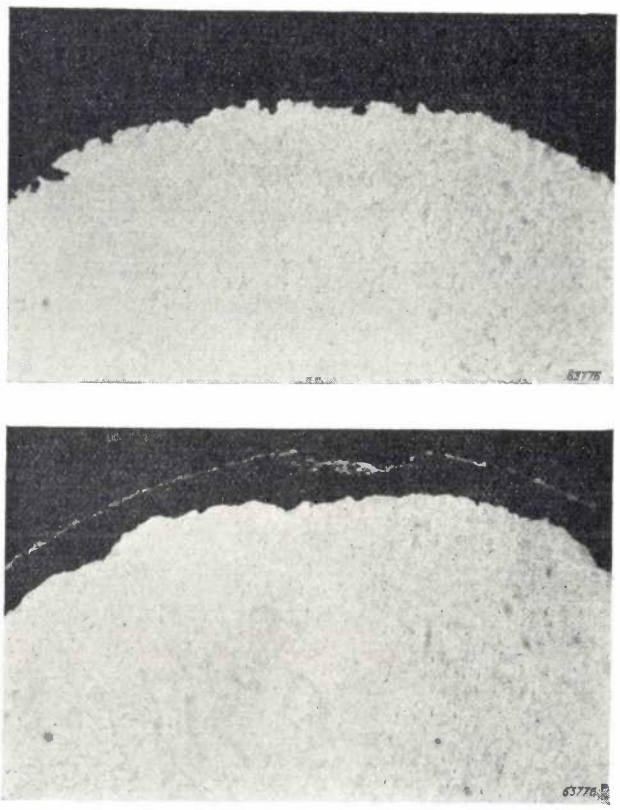


Fig. 5. *a*) Microphotograph of a specimen of the weakly etched cross section of a wire of AlCu 8/92 after heating for half an hour at 510 °C in MoO_3 and air. The black surrounding is synthetic resin ("Philite") which was pressed round the wire to facilitate flat grinding of the surface of the cross section. The corrosion on the surface of the wire is strongly localized (notch effect in the tensile test). Enlargement 600 times.
b) The same as *a*) but after three hours' heating of the wire. The corrosion here is decidedly more uniform. In the case of this specimen there is a light ring round the outside of the wire. In this layer, as revealed by X-ray diffraction examination, a layer of pure Cu has been formed which apparently has been precipitated from the liquid oxide mixture during the cooling of the wire after heating. Underneath this layer is a darker ring consisting mainly of molybdenum oxides (again found from X-ray diffraction). These oxides are responsible for the further rapid oxidation of the bare wire transferred, after removal of the loose oxide skin, to an atmosphere of pure air and further heated (cf. fig. 1).

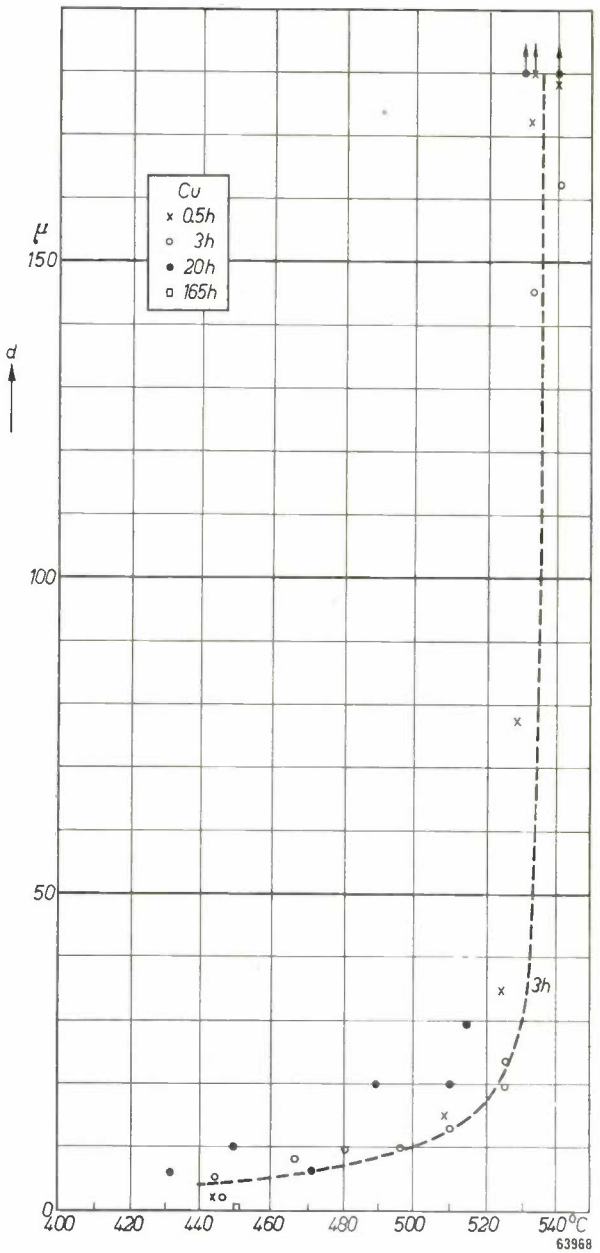


Fig. 6. Penetration depth as a function of the heating temperature in the case of the corrosion of wires of pure copper (OFHC) in MoO_3 and air.

[9]) Owing to droplets flowing together and forming larger drops the supply of MoO_3 may be locally interrupted, the corrosion then being localised to such an extent that it becomes very irregular.

Also with heat-resistant alloys of iron the simple case occurs of considerably increased oxidation taking place at the temperature at which the protective oxide itself reacts with the added oxide under the formation of liquid. In *fig. 7* the penetration depths have been plotted for chromium-iron CrFe 25/75. The greatly increased oxidation above 770 °C corresponds to the liquefaction of the eutectic MoO_3-MoO_2-Cr_2O_3. Of all the possible eutectics in the whole system this is not the lowest melting one: according to our measurements the

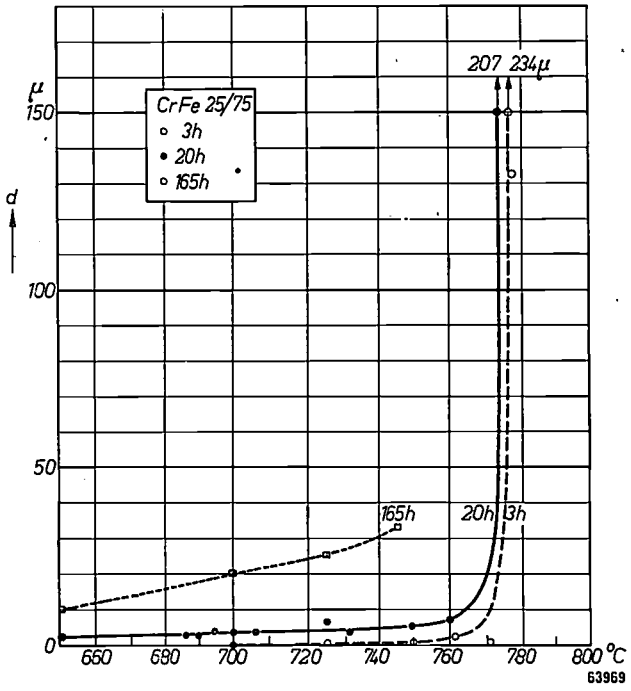


Fig. 7. Penetration depth as a function of the heating temperature in the case of corrosion of wires of CrFe 25/75 in MoO and air.

lowest eutectic temperature, of FeO-MoO_3-MoO_2-Cr_2O_3, lies at about 705 °C. Why this eutectic is not manifest in the curves of fig. 7 cannot be explained with certainty, but in principle it might be due to the fact that there is only very little FeO present in the protecting oxide skin, too little, in the case of reaction with MoO_3-MoO_2, to reach the domain of the (quaternary) phase diagram suitable for the formation of the lowest eutectic. There is also another low-melting eutectic which may occur in the said system of oxides and does not find expression in fig. 7, namely that with a temperature of about 725 °C. The influence of this eutectic, however, is quite clearly seen in the case of a chromium-iron alloy containing 9% nickel; see *fig. 8*.

Other factors tending to accelerate oxidation

On the ground of the agreement between a eutectic temperature and a discontinuity in the relation of penetration depth and temperature, the formation of liquid in the oxide skin is to be regarded as the cause of the accelerated oxidation observed. The formation of liquid, however, cannot be the only cause. Figures 7 and 8, for instance, clearly show that the presence of MoO_3 also causes a more or less gradual increase of oxidation of chromium-iron alloys below 705 °C, which in this case, as already mentioned, was the lowest eutectic temperature.

Now an acceleration of oxidation is to be expected also in the absence of liquid if molybdenum ions from the molybdenum oxide were taken up in the lattice of the protective oxide skin. It may be imagined that a lattice containing Mo-ions will not be stoichiometric, since Mo may quite easily occur in different valencies side by side. This means to say that in the lattice there will be unoccupied places making a relatively easy transportation of metal- or oxygen-ions possible, whilst on the other hand owing to an interchange of molybdenum-ions of different valencies a high order of mobility of electrons is ensured. Since, according to the theoretical views, obtained particularly by the work done by Wagner [10]) and by Mott [11]), in the oxidation of metals and alloys it is to be presumed that there is a separate diffusion of metal-ions and electrons through the oxide skin, in that skin with non-stoichiometric lattice we see literally a path free for the rapid oxidation of the underlying metal [12]).

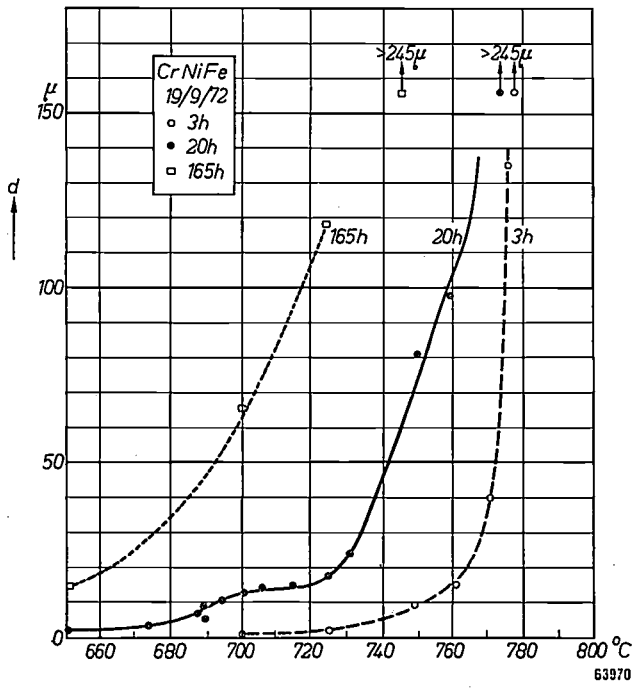


Fig. 8. Penetration depth as a function of the heating temperature in the case of corrosion of wires of CrNiFe 19/9/72 (stainless chromium-nickel steel) in MoO_3 and air.

[10]) C. Wagner, Z. phys. Chem. B **21**, 25, 1933; **32**, 447, 1936.
[11]) N. F. Mott, in N. Cabrera and N. F. Mott, Rep. on Progr. in Phys. **12**, 163, 1949.
[12]) Attention is drawn, in passing, to the parallelism between good protective action and poor electric conductivity of the oxide skin; for the latter point see Verwey, Haayman and Romeyn, Philips Techn. Rev. 9, 239, 1947/1948.

Absorption of a small percentage of molybdenum-ions in the lattice of the very thin oxide skin would be very difficult to prove experimentally. In the case considered here, however, the change brought about by the presence of MoO_3 appeared to be far more drastic: X-ray diffractional examination of

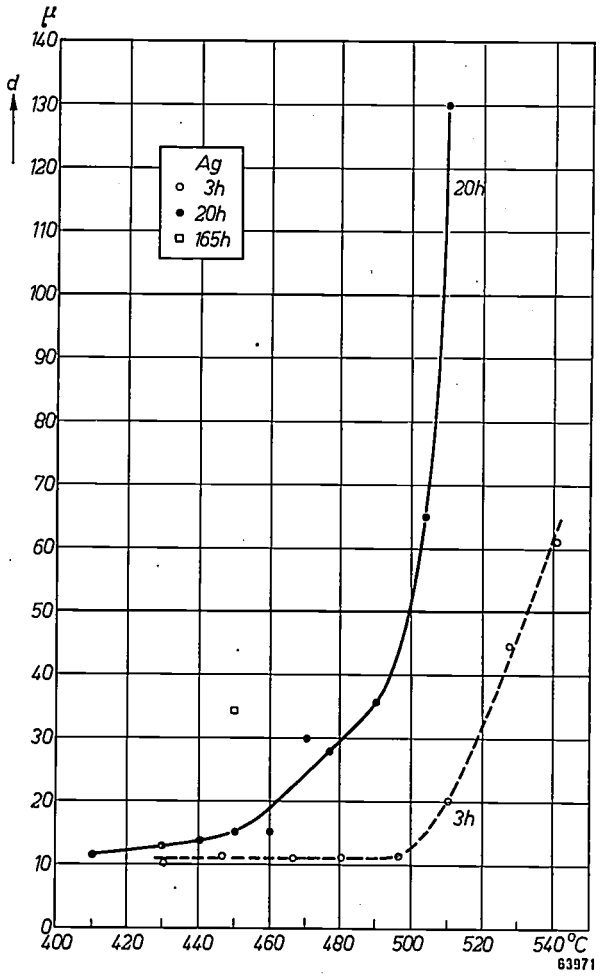


Fig. 9. Penetration depth as a function of the heating temperature in the case of corrosion of 99.6% pure Ag in MoO_3 and air. The silver wires were first preheated for 12 minutes at 600 °C in a dry mixture of nitrogen and hydrogen entirely free of oxygen.

the material in the oxide skin of the chromium-iron alloy CrFe 25/75 showed that after 165 hours' heating in MoO_3 and air at 650 °C there was no longer any sign of the structures normally to be expected (Cr_2O_3 and $FeO \cdot Cr_2O_3$), nor of Fe_3O_4, FeO and suchlike. Exclusively a new structure was found, the lattice type of which has not yet been ascertained but which, judging from a spectroscopic analysis, belongs to an oxide consisting for the greater part of molybdenum oxides (12% FeO or Fe_2O_3, 6% Cr_2O_3 and the rest Mo-oxides). Owing to its relatively high permeability, what there is of oxide skin cannot, therefore, afford any effective protection to the metal.

Yet another mechanism has to be assumed to account for our findings in the case of oxidation of silver and silver alloys. Experiments with samples of silver packed in MoO_3 powder showed at about 300 °C a noticeable depth of penetration increasing gradually up to 490 °C and then more pronounced at still higher temperatures; see *fig. 9*. At first sight the corrosion is astonishing, because, as mentioned in the introduction, above 150 °C in air of normal pressure silver cannot oxidize at all on thermodynamic grounds, owing to the great dissociation pressure of Ag_2O. In the presence of MoO_3 however double oxides are formed, such as Ag_2MoO_4, and these have a much smaller dissociation pressure [13]). We found that, even at its melting point of 570 °C, Ag_2MoO_4 does not decompose in air. Therefore, in the presence of MoO_3 silver behaves as a non-precious metal, which is only protected against rapid oxidation at high temperatures by an oxide skin. Therefore, we find in fig. 9, as in previous examples, the marked increase of oxidation at that temperature at which a eutectic in the oxide skin becomes liquid (495 °C is the lowest eutectic temperature in the system Ag_2MoO_4-MoO_3), and finally also strong oxidation in the case of contact only with MoO_3 vapour at temperatures where the vapour pressure of MoO_3 is not too low (750 °C). It is seen, therefore, that by the addition of MoO_3 to Ag_2O the temperature at which the dissociation pressure exceeds 1/5 atmosphere is raised by at least 600 °C. It is possible that at temperatures higher than 750 °C, to which we did not go in our experiments, there is no longer any oxidation of silver in air containing MoO_3.

Although this makes the oxidation of silver by MoO_3 at high temperatures understandable, there is a remarkable feature about this oxidation that needs explaining, namely the small temperature-dependency of the measured penetration depth below the lowest eutectic temperature. When, instead of the penetration depth, we measure the loss in weight as function of the temperature, then the temperature-dependency below the lowest eutectic temperature appears to be much greater. This would indicate that the "temperature-insensitive" corrosion below 490 °C must, at least for a large part, be due to intercrystalline penetrations, which at increasing temperature do not increase so much in depth as in number or in width. Such penetrations have indeed been found microscopically in specimens of wires heated below 490 °C. (Above 490 °C the penetrations are still more pronounced; see *fig. 10*.)

Such a penetration, at temperatures where no liquid occurs and where the vapour pressure of MoO_3 is entirely negligible,

[13]) For analogous double oxides, e.g. of silver with vanadium or with chromium, something similar has been known for a long time: see R. Schenck, Z. anorg. allg. Chem. 249, 88, 1942.

is difficult to understand. It may be that owing to the reaction of Ag and MoO_3 in the solid state, and the attendant changes in volume, strong mechanical stresses arise in the silver crystals and cause fractures along the grain boundaries. Such

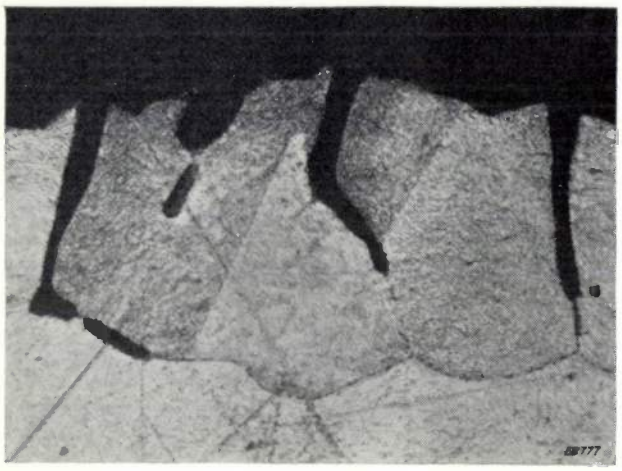


Fig. 10. Microphotograph of a specimen of the cross section of a block of silver (about 99.5% Ag) oxidized by heating for 20 hours at 530 °C in MoO_3 and air. There is strong intercrystalline corrosion, at least along the boundaries of grains orientated at random. On the straight boundaries separating twin parts of a grain the crystalline structures to the right and left are exactly matched. The interfacial energy on these boundaries is thus very small, so that the liquid (formed at 495 °C) does not penetrate there. Also these twin boundaries have been made clearly visible by etching the specimen with $NH_4OH + H_2O_2$. Enlargement 500 times.

a corrosive mechanism would have to be rather independent of temperature, but it would be strongly dependent upon small contaminations at the grain boundaries. Probably there is a relation between this and the fact that in the case of silver this corrosion appeared to be the more intense the less the degree of purity of the silver.

The three effects discussed here: formation of a liquid, formation of a non-stoichiometric oxide lattice, and displacement of a thermodynamic dissociation equilibrium, probably play a part in many hitherto unaccountable cases of corrosion of heat-resistant alloys in the presence of foreign substances. Though there may, of course, be still other important mechanisms at work, in all the cases of oxidation of alloys investigated by the authors the three mechanisms mentioned were sufficient to account satisfactorily, on broad lines, for the phenomena observed.

Summary. Alloys which are protected against corrosion by a dense skin of oxide, and thus may serve as heat-resistant constructional materials, sometimes show a strongly accelerated oxidation when brought into contact with foreign substances, especially foreign oxides. With a view to accounting for this phenomenon, the oxidation of a series of metals and alloys, i.a. with copper, iron and silver as base, has been investigated for the case where these are brought into contact with molybdenum oxide (MoO_3 in powder form). The degree of oxidation ("penetration depth") was determined by measuring the breaking strength of wires drawn from the respective material and heated in air with and without contact with MoO_3. When the oxidation is thus determined as a function of temperature a curve is obtained showing one or more discontinuities where the oxidation is suddenly accelerated. These discontinuities appear to lie at the eutectic temperatures of the systems of the respective oxides (lowest temperatures at which the formation of liquid occurs). From this it can be concluded that the accelerated oxidation is due to formation of liquid in the protective oxide skin of the alloy. The accelerating effect of the liquid upon oxidation is readily understood, and from the relation that has been found it is also possible to explain many peculiarities in the temperature- and time-dependency of the oxidation. In some cases oxides of the basic metal not essentially belonging to the protecting skin, such as Cu_2O in the case of aluminium bronze, may considerably contribute towards a lowering of the eutectic temperature. Sometimes in the presence of MoO_3 perceptible oxidation already takes place below the lowest eutectic temperature. This is accounted for by the absorption of molybdenum ions in the crystal lattice of the oxide skin and the accompanying greater freedom of movement of electrons and metal- or oxygen-ions. A third way in which MoO_3 may promote oxidation is found in the case of silver. Silver, as a precious metal, is not oxidizable above 150 °C in air of normal pressure, but in the presence of MoO_3 double oxides may be formed, making it possible for oxidation to take place at much higher temperatures; then again melting phenomena in the oxide skin make their influence felt.

ON THE REGULATION OF INDUSTRIAL PROCESSES

by H. J. ROOSDORP.

621-52

Regulating systems for industrial processes may be of a widely different nature: think, for instance, of the governor used on the oldest steam engines for keeping their speed constant, the voltage and frequency stabilizers for generators, the temperature regulators that are so highly important in the chemical industry, for distilling and cracking in the oil industry and for the preserving of foodstuffs. All these apparently quite different devices have one principle in common, that whatever it is that is to be kept constant is fed back via a controller upon a controlling quantity governing the first mentioned quantity.

It is the intention to follow up this article, giving a general review of such regulating systems, by others dealing in turn with a kind of electronic calculating machine for solving regulating problems and with a developed regulating device.

For most industrial processes it is necessary that some physical quantity, such as temperature, pressure, rate of flow, acidity or conductivity, should be kept at a certain level or value within a given tolerance. For this purpose there must be some regulating unit upon the position of which the quantity to be kept constant depends, as for example the gas valve of a gas-fired furnace in which the temperature must be kept constant. In the simplest cases it suffices to set the regulating unit in a certain position derived from theoretical calculations or practical experience (*fig. 1*). Allowance may be made as far as possible for the variations to be expected in all factors influencing the quantity to be regulated. In the case, for instance, of the furnace just mentioned, such factors are

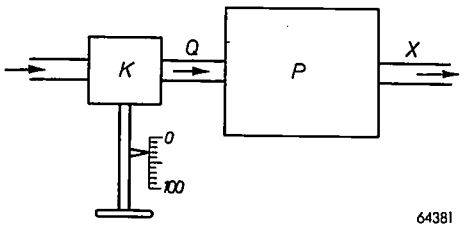


Fig. 1. To bring, for instance, a gas-fired furnace P to a certain temperature, with a valve K the gas flow Q can be pre-adjusted once for all so that the temperature X assumes and maintains as well as possible the desired value. Here there is no influence of the output quantity X upon the input quantity Q (open circuit).

the charge in the furnace, the pressure of the gas, the ambient temperature, etc. During the "process" — in this case the generation of heat inside the furnace — there is no feedback of the temperature in the furnace upon the position of the gas valve. There is no provision for correcting, while the "process" is proceeding, any unexpected changes subsequently taking place in the working conditions and affecting the temperature. Regulating — if it may still be called that — is therefore limited, in such an open circuit, to choosing a predetermined position of the regulating unit as well as possible.

A closed circuit (*fig. 2*) has a different behaviour. Here the quantity to be regulated is continuously measured and from information thus obtained the adjustment of the regulating unit to reach or maintain the conditions desired is deduced. Henceforth regulation is to be understood as referring only to this system. In this case there is a continuous influence (feedback) — making the circuit a closed one — of the output of the process upon the input. This feedback may be an observer (manual regulating system, fig. 2*a*) or an automatic controller (automatic regulating system, fig. 2*b*).

In what follows we shall only consider closed circuits with automatic controllers.

Accuracy and stability

The ideal regulating system would be that which works in such a way that the quantity to be regulated can never deviate from the desired value in spite of interfering, mostly random, variations in all sorts of working conditions which, without regulating system, would have a prohibitive effect upon the process. When it is desired to change over to some other value of the quantity to be regulated some adjustment has to be made to the controlling unit, and an ideal regulating system must also immediately respond to the adjustment.

It is due to the limited sensitivity and the inertia of various elements of the circuit that an actual regulating device does not fully answer the ideal in either of these two respects. By making the

controlling unit highly sensitive it is of course possible to arrange that the slightest deviation from the desired value causes a considerable change in the position of the regulating unit, but owing to the time delay of the detecting element, of the controller and of the regulating unit, the latter

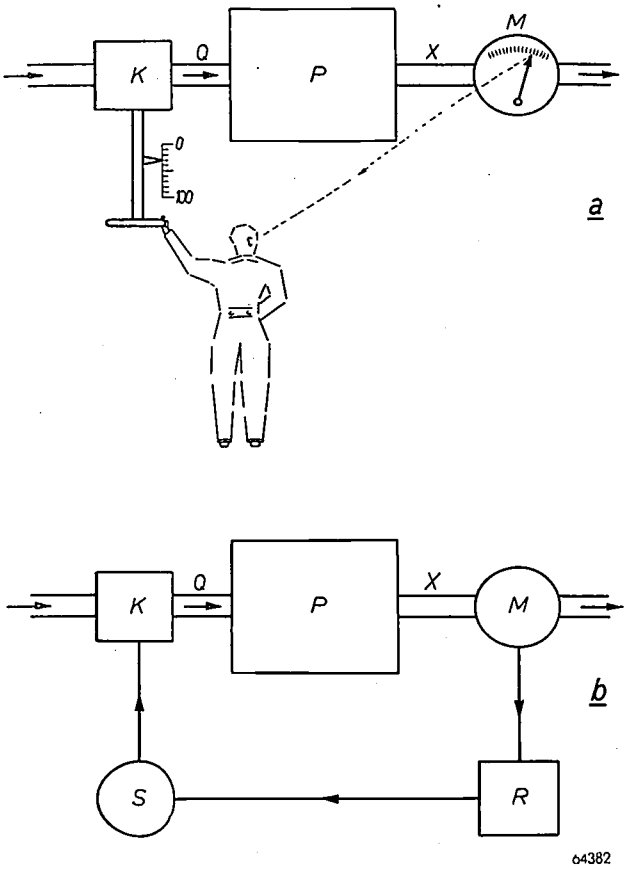


Fig. 2. Closed regulating circuits. The output quantity X to be regulated is measured by a detecting and measuring element M. According to the magnitude of the deviations from the desired value of X, the input quantity Q is adjusted, by means of the regulating unit K, through the intermediary of either (a) an observer or (b) an automatic controller R and a servomotor S.

does not reach its new position until some time after the disturbance has taken place, and in consequence of the (often very considerable) inertia of the process the effect of the said adjustment is delayed.

After a disturbance — at least temporarily — there is consequently a certain difference between the actual and the desired value of the quantity to be regulated. When a new stationary state has been reached this deviation may in some cases be zero or the difference between the two values mentioned may be reduced to within certain limits depending, inter alia, upon the sensitivity of the detecting element (which may usually be ignored), but there are other cases where the deviation may remain at a certain value, and one then speaks of static inaccuracy, or offset.

The deviation which, as already said, is always present prior to the stationary state having been reached is called the transient phenomenon. Here there are three cases to be distinguished: (1) the aperiodic case, where the deviation always has the same sign while gradually decreasing, (2) the damped periodic case, where the deviation diminishes as the decay of a damped oscillation, and (3) the non-damped periodic case, when the deviation continues to fluctuate between certain limits without decaying.

The importance of offset is immediately obvious: it denotes in how far the quantity to be regulated approaches (in course of time) the desired value when a disturbance has taken place. But also the dynamic behaviour of the system is important: the deviations from the desired value following upon a disturbance might, for instance, lead to an intolerably large amplitude, or it might take far too long before a new stationary state is reached. It may even happen that the desired value is not reached at all owing to the input quantity undergoing every time another disturbance before any stationary state can be attained, each of these new disturbances affecting the output to be regulated continuously. As an example may be mentioned the regulating of the pressure in a drum filled with gas supplied by a piston pump. In such a case there is a strong pulsatory variation in the gas pressure following the frequency of the pump. This fluctuation, the amplitude of which may possibly reach intolerable proportions, is called the dynamic inaccuracy.

Aperiodic and damped periodic cases are said to be stable (since a stationary state is reached by the time a new disturbance arrives), while non-damped periodic cases — also referred to as hunting — are unstable. Farther on we shall see that a certain group of controllers (the so-called discontinuously acting controllers) are unstable by their very nature, so that there is always inevitably a certain fluctuation in the quantity to be regulated. If such a fluctuation is to be regarded as prohibitive a so-called continuously acting controller has to be used, which may under certain conditions be stable.

The two kinds of inaccuracy, the stability (in the case of continuously acting controllers) and the transient are the main factors to which particular attention must be paid in the regulating technique. They depend, of course, upon the properties of

the elements comprising the closed circuit. Usually the plant is already defined and the regulating system has to be adapted to it in such a way that the desired accuracy and stability are obtained. If the requirements were to be worked out entirely by calculation then in many cases insurmountable difficulties would be encountered. On the other hand it is not advisable either to try to find the best regulating system solely by empirical means during the manufacturing process, because this takes much time and might upset the production for long periods. What has proved to be the best method in many cases is to study the situation with the aid of an electrical model.

Working with an electrical model

When working with a model the measurements to be taken from the plant itself can be limited to the recording of a step-function response curve representing the variations of the output quantity (which is to be regulated) as a function of time when the input is changed according to a certain function of time, for instance as a sudden change. Then an electrical network must be designed having a response curve, on a different time scale, corresponding as closely as possible to that of the process. This network then serves as a model, the behaviour of which, in combination with a model of the automatic controller, can be studied quickly and without interfering with the actual process. Guided by some experience one chooses for the model of the automatic controller one of the various types that are possible.

As a rule there is a considerable difference between the time scale of the actual process and that of the electrical network: a change requiring several minutes, for instance, in the process may take only a few milliseconds in the network. A periodic repetition of the sudden change makes it possible to observe the response curve of the network, and also the behaviour of the model with the automatic controller on the screen of an oscilloscope. An apparatus has been developed, called the electro-analogue, in which are incorporated all the elements required for building up any desired model of the process and of the automatic controller, as also an oscilloscope, a generator supplying voltage pulses ("step functions of voltage") and other accessories. Further articles will be devoted to the description of this electro-analogue.

Once the parameters of the model of the automatic controller have been chosen for optimum regulating conditions, then — taking into account the ratio of the time scales and the sensitivities — it is possible to derive the values of the settings in the actual regulating device.

Another method of determining the behaviour of a process is to plot the frequency response curve. To this end, the input quantity is caused to fluctuate sinusoidally with a constant amplitude about a normal value, at different frequencies. Measurements are then taken of the amplitude of the output quantity (in relation to the amplitude of the input) and of the phase shifts taking place between the fluctuations of the input and the output quantities. These two measured quantities can in our case most advantageously be worked out in a polar diagram. With this curve, together with the corresponding curve for the automatic controller, the stability of the regulating system can be judged.

The step-function response curve gives theoretically just as much information as the frequency response curve, with the advantage that it can be recorded much quicker and more easily than the frequency response curve. Taking again the example of a gas-fired furnace, the step-function response curve is obtained by suddenly opening the gas valve a little wider or turning it back a little and then measuring from that moment onwards the temperature inside the oven as a function of time. To record a frequency response curve it would be necessary to vary the gas supply sinusoidally and at a variable frequency and then, after a stationary state has been reached, the amplitude of the temperature fluctuation and the phase difference with the fluctuating gas supply have to be measured.

A second advantage of the method of working with step-function response curves is the simple way of designing the model in which the behaviour of the process is to be imitated. When the step-function response curve of this model is displayed on the screen of an oscilloscope it is not difficult to make it equivalent to that of the process by varying some of the network parameters. It would not be so easy to find agreement of the frequency response curves by oscilloscopic means.

Against these advantages of the step-function response curve, however, is the objection that disturbances occurring in actual processes are not usually of the nature of an impulse but mostly more or less gradual and periodical, so that the plotting of the frequency response curve is more in line with actual conditions. Furthermore, natural oscillations of the process often find better expression in the frequency response curve than in the step-function response curve.

While plotting either of these curves in the process it is necessary to see that all other working

conditions — in the case of a furnace, for instance, the charge, the ambient temperature, etc. — are kept constant.

In the foregoing it has been tacitly assumed that the amplitude of the input disturbance or the magnitude of the step function is immaterial, and that the process is therefore linear. Often, however, it is only in the case of small variations that the process can be regarded as linear with a sufficient approximation. A complete investigation would have to cover also the behaviour of the process in the event of larger amplitudes, and the non-linearity of the actual process would have to be imitated as closely as possible by using non-linear elements in the network.

Any deviation from the linearity can be judged firstly by varying sinusoidally the input quantity of the process and observing in how far the output quantity is non-sinusoidal, and secondly by determining in how far non-linearity exists between input disturbance and output for disturbances of different magnitude.

In these considerations, however, we shall disregard non-linear effects.

Finally it is possible, at least in theory, to treat the problem of regulating mathematically. To facilitate the formulation of the necessary equations the letter Q is taken to represent the position of the regulating unit and the letter X for the deflection of the measuring element indicating the regulated quantity. This will prove to be very practical also when recording the step-function or the frequency response curve.

The fact that this definition differs somewhat from that given for fig. 2 is understandable when one bears in mind the example of the gas-fired furnace. Owing to the time delay of the detecting element the reading of the measuring element (now denoted by X) will not entirely agree with the temperature prevailing at the very same moment, originally denoted by X, and the flow of gas (Q in fig. 2) depends, it is true, upon the position of the gas valve now denoted by Q, but it also depends upon such factors as the pressure in the supply line. In the new definition the detecting element and regulating unit are reckoned as belonging to the plant.

In the relation between Q and X describing the process there will, as a rule, be also differential quotients with respect to the time t:

$$f_1\left(Q, \frac{dQ}{dt}, \frac{d^2Q}{dt^2}, \ldots, X, \frac{dX}{dt}, \frac{d^2X}{dt^2}, \ldots\right) = 0. \quad (1)$$

(When Q is given a step-function or a sinusoidal variation with time, then it is possible to derive from this equation the step-function response curve or the frequency response curve as the case may be.) For the automatic controller, too, a relation between Q and X can be formulated:

$$f_2\left(Q, \frac{dQ}{dt}, \frac{d^2Q}{dt^2}, \ldots, X, \frac{dX}{dt}, \frac{d^2X}{dt^2}, \ldots\right) = 0. \quad (2)$$

When the plant and the automatic controller are joined together to form one closed circuit (fig. 2*b*) then equations (1) and (2) may be combined. From these equations and the initial conditions one can, at least in theory, solve X and Q, which are then found as functions of time:

$$X = f_3(t), \quad \ldots\ldots\ldots \quad (3)$$

$$Q = f_4(t) \ldots\ldots\ldots\ldots \quad (4)$$

From the curve representing (3) it can be ascertained whether the regulation is stable and whether it is sufficiently accurate, thus, e.g., whether X remains within the limits set or if, in the event of it exceeding those limits, it returns within them quick enough. From the graph representing (4) it can be seen whether the necessary variation of Q falls within the range covered by the regulating unit.

Actually, however, it is not as a rule an easy matter to solve X and Q from the equations (1) and (2). It is then the electrical model that provides the solution: with this model it is easy to observe the variations of the output and input voltages with time and to ascertain the effect of changes of certain parameters, introducing a disturbance at a suitable point in the circuit, either a step or a sinusoidal disturbance. In the latter case it is preferable to choose a frequency corresponding to the frequency of the disturbances occurring in practice, obviously taking into account the ratio of the time scales.

The quality of a regulating system can be judged from various points of view. Sometimes it suffices to consider the maximum deviation in the output quantity caused by a disturbance of a certain magnitude introduced somewhere in the system. In other cases the time taken for the output quantity to be brought back and kept within certain limits after the occurrence of a certain disturbance is more important; see *fig. 3a*. If a small deviation of long duration is in principle just as undesirable as a large deviation of short duration, then it may be taken as a criterion, for instance, that the hatched area in fig. 3*b*,

$$\int_0^\infty |X - X_0| \, dt,$$

in which X_0 represents the desired value of X, or the integral

$$\int_0^\infty (X - X_0)^2 \, dt,$$

has to be kept below a certain value. No appreciable

error is made if the integration is extended only over a time T of such a length that X becomes practically constant within that time; the values of the integrals extended over T can then easily be measured electrically in the model, the first value with the aid of a rectifier and a moving-coil meter and the second with a thermocouple and a moving-coil meter.

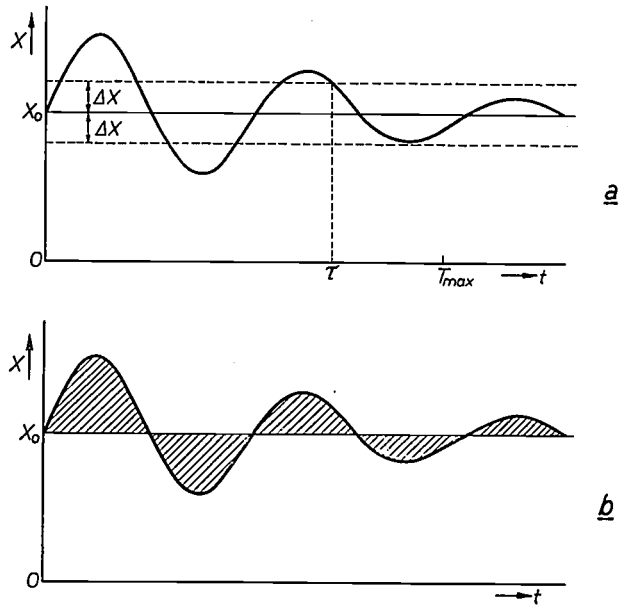


Fig. 3. Two criteria for the potential correction of a regulating system. In the case (*a*) the criterion is that the time t required for the quantity X to return permanently within the limits X_0-X and X_0+X must be less than a given time T_{max}. In the case (*b*) it is required that the hatched area does not exceed a certain value.

Types of regulating systems

Equation (2) expresses in what way the output of the controller depends on the input quantity. In principle one is quite free in the choice of this function, but in practice one is limited to functions that are easily attainable with mechanical or electrical switching elements. The most common combinations are mentioned below.

These are to be divided into two main groups: continuously and discontinuously acting controllers. The latter being the simpler in construction, for which reason they were first preferred in the technique of regulation (temperature control of furnaces, of centrally heated buildings, and suchlike), these will be dealt with first.

Discontinuously acting controllers

A controller is said to be discontinuously acting when the regulating unit takes up a different position discontinuously upon X passing a given critical value.

We shall first consider the case where all working conditions are normal. Let Q_0 be the value of Q necessary to give X the desired value X_0. In the simplest case the regulating unit is so arranged that the position Q_0 is not possible, there being only two possible positions: Q_1 ($< Q_0$) and Q_2 ($> Q_0$). According to the process equation, Q_2 corresponds to a value of X less than X_{cr}, while Q_2 applies for a value of X greater than X_{cr}. The controller reacts to the magnitude of X in such a way that either $X > X_{cr}$, $Q = Q_1$, or $X < X_{cr}$, $Q = Q_2$. One then speaks of a regulating system with two-step action (*fig. 4*). Naturally such a system is always unstable, for when X, for instance, is too great ($X > X_{cr}$) the regulating unit assumes the position $Q_1 < Q_0$ and Q therefore becomes too small, so that after a time X also becomes too small, with the result that the regulating unit again takes up the position Q_2 and X becomes greater again, and so on. Thus X fluctuates continuously about the value X_{cr} and the regulating unit alternates between the positions Q_1 and Q_2. Generally one will try to make the mean value of X equal the desired value X_0. In a certain working position X_{cr} may be equal to X_0 (see the part of *fig. 5* between $t = 0$ and $t = t_0$).

Let us now take a case where one of the working conditions changes, or where the controller is so adjusted that the regulating unit changes position at a different critical value X_{cr}' (see fig. 5 to the right of t_0). The only manner in which the controller can respond to the new situation is by changing the ratio of the intervals in which the regulating unit takes up the two possible positions. Since X_{cr}' is assumed to be greater than the original value X_{cr} the regulating unit has to remain

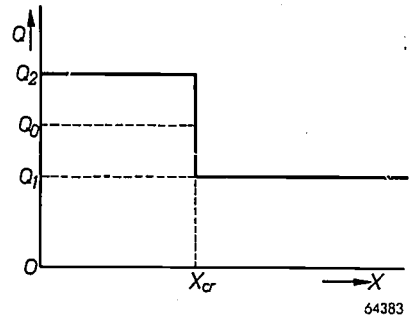


Fig. 4. Regulating system with two-step action. When X is greater than a certain critical value X_{cr} then Q has the valuc Q_1, and when X is less than X_{cr} then $Q = Q_2$.

longer in the high position Q_2 than in the low position Q_1. The new value X_0' around which X now fluctuates is higher but, as the diagram shows, remains lower than the new adjustment of the controller, X_{cr}'.

A regulating system with two-step action can be extended to one with multi-step action working according to the following scheme:

when $X > X_2$ then $Q = Q_1$,
when $X_1 < X < X_2$ then $Q = Q_2$ $(>Q_1)$,
when $X < X_1$ then $Q = Q_3$ $(>Q_2)$.

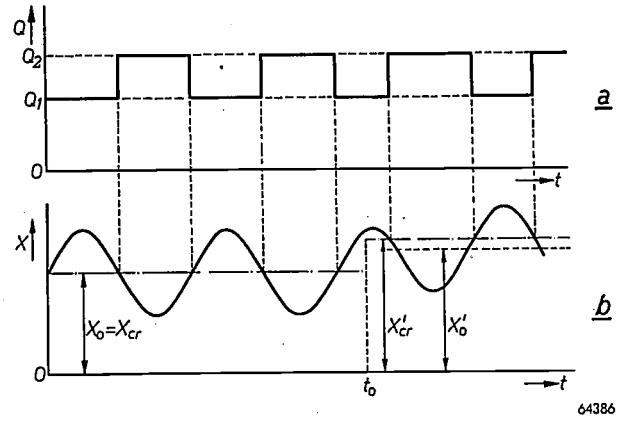


Fig. 5. Variations of Q and X in a regulating system with two-step action. Prior to t_0, X_{cr} was equal to the desired value X_0, around which X fluctuates, and Q had alternately, during equally long intervals, the values Q_1 and Q_0. At t_0 X_{cr} is raised to the value X_{cr}', as a consequence of which Q has each time the value Q_2 for a longer interval than the value Q_1. The value X_0' around which X then fluctuates is less than X_{cr}'.

Such a system has the advantage of reducing the variations of the input quantity and also the fluctuations of the process, so that the regulating unit remains at rest during longer intervals, thus being subject to less wear than in the case of a regulating system with two-step action.

Instead of causing the quantity Q itself to change at certain values of X, also the rate at which Q changes, thus dQ/dt, can be made to assume a different value. One then speaks of a multi-speed floating system. *Fig. 6* is a representation of the case where there are two speeds, p_1 and p_2. In practice there always is a certain zone (dead zone) between p_1 and p_2 where the system is at rest:

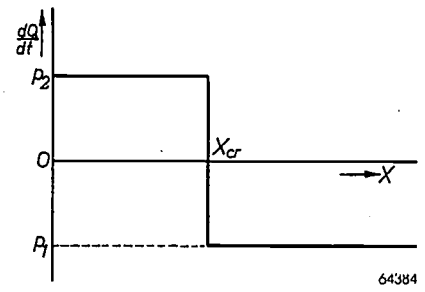


Fig. 6. Multi-speed floating system. Upon X passing the critical value X_{cr} the speed dQ/dt with which the regulating unit is adjusted suddenly changes from the value p_1 to the value p_2, or vice versa. p_1 and p_2 have opposite signs.

when $X > X_2$ then $dQ/dt = p_1 < 0$,
when $X_1 < X < X_2$ then $dQ/dt = 0$,
when $X < X_1$ then $dQ/dt = p_2 > 0$.

This system of regulation is a transition to that of continuously acting regulating systems, since only the speed changes discontinuously, whereas the position of the regulating unit changes continuously.

Continuously acting controllers

The simplest form of a continuously acting regulating system is that with proportional action where equation (2) takes the simple form of:

$$Q = a(X - X_0) + Q_0, \quad \ldots\ldots \quad (5)$$

in which a is a constant. Q_0 is the position of the regulating unit for a stationary state of $X = X_0$ (eq. (1)).

In the event of a disturbance, caused by a change in the process equation, as a rule Q will have to assume a different value if the process is again to be brought to the state where $X = X_0$. Q and $a(X - X_0)$ — if Q_0 is left unchanged — can only vary by the same amount, and thus in the new position X will differ from X_0. This difference is the offsett of the regulating system. The greater the value of a, the sensitivity of the controller, the less is this difference, but the increasing of a is limited by the stability in the same way as the amplification of an amplifier with positive feedback: when a certain value is exceeded oscillations arise, that is to say, the system becomes unstable.

This offset is avoided in a controller with integral action. Between Q and X we then have for the controller an equation of the form:

$$Q = b\int (X - X_0)\, dt + Q_0, \quad \ldots \quad (6a)$$

hence

$$\frac{dQ}{dt} = b\,(X - X_0), \quad \ldots\ldots\ldots \quad (6b)$$

in which b is a constant. In this case, therefore, the speed at which the regulating unit is adjusted is proportional to the deviation of the position of the measuring element from the desired value X_0. This system is therefore also called a proportional-speed floating system. So long as this deviation is not reduced to zero, the factor $\int (X - X_0)\, dt$ continues to increase, the regulating unit thus continuing to move in the direction for correcting the deviation, until ultimately $X = X_0$. Consequently there is then no offset. The greater the value of b the quicker the final state $(X = X_0)$ is reached, but here again a limit is set to the increasing of this coefficient owing to instability setting in.

Combinations of controllers with proportional and integral action (compound action) are also possible:

$$Q = a(X - X_0) + b \int (X - X_0)\mathrm{d}t + Q_0,$$

and there are also combinations of each of these systems, or of both, with a third type: the controller with derivative action, where a term $c\,\mathrm{d}X/\mathrm{d}t$ has to be added to the relation between Q and X in the right-hand member. This constitutes a contribution towards the adjustment of the regulating unit which is proportional to the rate at which X departs from the desired value (or is approaching that value). As a rule it brings about a quicker approach to the final state, since a and/or b can be chosen larger without risk of the regulating system becoming unstable.

In principle it should be possible to construct regulating systems with a relation between Q and X containing also higher derivatives or multiple integrals of X, but for industrial purposes there seems to be little need of these in practice.

It is to be noted that integrating and differentiating can be carried out electrically to a very good approximation, as is in fact done in the electrical model of a continuously acting controller. In many cases it is advantageous to construct also the controller itself as an electrical apparatus, as will be shown in a subsequent article following the two next ones devoted to the electro-analogue, in which the functioning of different regulating systems will be explained with the aid of oscillograms.

Summary. In this introductory article on regulating systems for industrial processes emphasis is laid upon the offset and the dynamic inaccuracy, the stability and the damping, as being the most important points to be considered in the designing of a regulating device. Attention is drawn to the advantages of working with an electrical model of the process to be regulated and of the controller to be designed for it; with this model the phenomena can be observed with the aid of an oscilloscope. The controllers are classified under discontinuously and continuously acting controllers. In the latter it is possible to introduce, in different ratios, a proportional, an integral and a derivative action.

Further articles will be devoted to a discussion of the "electro-analogue" — a combination of instruments with which analogous models of any processes to be regulated, including the controllers, can be built up — and a description of the actual construction of a controller.

PRECISE CALIBRATION OF TUNING FORKS

by C. C. J. ADDINK.

681.831.3:534.321.71:534.63

The introduction into the field of music of the new international standard of concert pitch based on a frequency of 440 c/s for the note A_3 in the middle octave instead of 435 c/s still involves difficulties owing to the lack of good standards, the tuning forks used in practice leaving much to be desired. Hence the greater interest now being shown in accurately calibrated tuning forks.

An international conference held in London in 1939 recommended that such bodies as the "Comité Consultatif International Téléphonique", the "International Electrotechnical Commission" and the "Union Internationale de Radiodiffusion", as also the various national standardisation committees, should fix the concert pitch at a frequency of 440 c/s for the note A_3 of the middle octave — instead of the 435 c/s internationally accepted at Vienna in 1885 — and that this should be adhered to as closely as possible in all musical performances[1]). In practice the usual tuning forks are, it is true, tuned higher than 435 c/s but often differ, in the positive or the negative sense, several c/s from the figure of 440 c/s. Since there was no standard for the frequency of 440 c/s in this laboratory a method has been developed, as the first step in this direction, for measuring audio frequencies direct (i.e. by counting the number of oscillations in a time interval) and with extreme accuracy. With the aid of an apparatus working according to this method not only have tuning forks of 440 c/s been calibrated but also sets of 13 tuning forks, forming a tempered chromatic scale of one octave (*fig. 1*). The method referred to will be described here, but the question to be considered first is the degree of accuracy to be demanded from a tuning fork.

[1]) See, e.g., Internationale Neufestsetzung des Stimmtones, Akustische Zeitschrift **4**, 288, 1939.

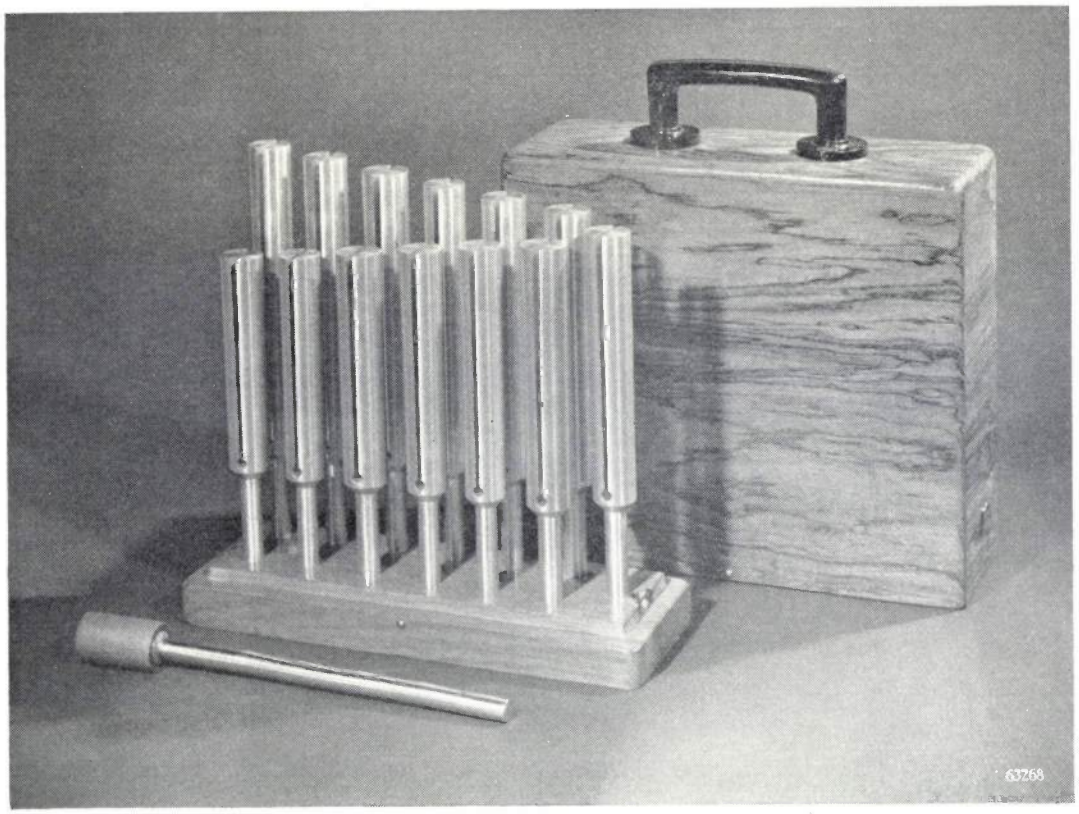


Fig. 1. Set of 13 tuning forks made in the Philips Laboratory. Together they form a tempered chromatic scale from c^1 to c^2, with $a^1 = 440$ c/s. The reason for the cylindrical shape is explained in the last paragraph.

Accuracy required of a tuning fork

The useful decay time of a normal tuning fork is 5 to 15 seconds. When the pitch of a musical instrument is compared with the tuning fork in the normal way one can usually hear beats. If, however, the difference in frequency is so small that there are no more than two or three beats during the decay time then they are hardly noticeable as such and one has reached the limit of the accuracy of the tuning fork. It serves no purpose, therefore, to require any greater precision than 0.2 c/s (one beat in 5 seconds), or in other words the frequency of the tuning fork has to lie between the limits of 439.8 and 440.2 c/s [2]). (This degree of accuracy is obviously only required for musical instruments with a fixed tuning, such as the organ, the piano, the harp, etc.) The frequency, however, has to be determined with greater accuracy, because one must be quite sure of the last decimal in the said limits. The permissible measuring error is therefore 0.05 c/s, which at a frequency of 440 c/s amounts to a measuring accuracy of about 0.01%.

This implies that account must be taken of the temperature when measuring and when using the tuning fork in practice. The temperature coefficient of the frequency in the case of a steel tuning fork is in the order of -1×10^{-4} per °C, while for the forks made of a hard aluminium alloy illustrated in fig. 1 it is about -2.5×10^{-4} per °C.

Method of measuring

As is evident from the foregoing, for calibrating a tuning fork a more accurate method is needed than that where the tuning fork is compared with a standard frequency by ear. Furthermore, with frequency differences less than 0.1% the sign of the error can no longer be determined by ear.

The method of calibrating worked out by us briefly amounts to the following. The frequency of an auxiliary oscillator is matched with that of the tuning fork to be calibrated and then a synchronous clock is connected to the oscillator and the oscillations made by the auxiliary oscillator in a certain interval of time are counted. This time interval is measured with the aid of a calibrated pendulum timepiece.

Matching the auxiliary oscillator

Owing to the damped character of the vibration a tuning fork is not suitable for measuring the frequency with sufficient accuracy without some other aids. It is true that a tuning fork can be kept in vibration by electrical means, but then, unless a great deal of care is taken, the frequency at which it vibrates differs somewhat from the frequency of a fork vibrating freely, and it is just this free vibration that has to be measured accurately. What is needed for calibrating, therefore, is an auxiliary oscillator which, once it has been matched with the freely vibrating fork, continues to oscillate in that frequency for an unlimited length of time.

An *RC* oscillator consisting only of resistors, capacitors and amplifying valves is very suitable for this purpose. If such an oscillator is properly designed, when it has reached temperature equilibrium the relative frequency change in the audio-frequency range can be kept smaller than $1:10^5$ [3]).

The matching of the frequency of the oscillator with that of the tuning fork can be done with sufficient accuracy if a cathode-ray oscilloscope is employed, by applying the signal from the *RC* oscillator to one pair of plates and a signal with the frequency of the tuning fork to the other pair. The latter signal can be obtained by setting up the tuning fork in front of a microphone. There then appears on the screen of the oscilloscope a changing Lissajous figure which, when the frequencies are equal, becomes a stationary ellipse [4]).

There is, however, a drawback to this method. The Lissajous figure cannot demonstrate whether the frequency deviation is positive or negative, hence the operator does not know in which direction to turn the oscillator tuning knob. If this is turned in the wrong direction the consequent time loss results in the tuning fork signal having decayed too far, and the fork has to be struck again.

To obviate this, the following method is adopted. The signal from an *RC* oscillator is applied direct to one pair of input terminals of the oscilloscope and also, after having been shifted 90° in phase in a network of resistors and capacitors, to the other pair of terminals (*fig. 2*). Given the right proportions of amplitude of the two signals, a circle is then produced on the screen of the oscilloscope.

The tuning fork to be calibrated is set up front in of a microphone, the amplified signal from which

---

[2]) Other authors arrive at the same tolerance another way: H. J. von Braunmühl and O. Schubert, Ein neuer elektrischer Stimmtongeber für 440 Hz, Akustische Zeitschrift **6**, 299-303, 1941.

[3]) See, e.g., E. L. Ginzton and L. M. Hollingworth, Phase-shift oscillators, Proc. Inst. Rad. Engrs **29**, 43-49, 1941; K. Bucher, *RC*-Generatoren, Telegr. Fernsprech-Techn. **31**, 307-313, 1942.

[4]) For this and the other method to be described later it is an advantage to use an oscilloscope with two amplifiers, one for each pair of plates, such as type GM 3159 or GM 5655 (Philips Techn. Rev. **9**, 202-210, 1947, and **11**, 111-115, 1949 (No. 4)).

is used for modulating the beam current of the cathode-ray tube, such that the beam current is allowed to pass only during the positive half cycles of the microphone signal. Thus only half a circle is seen on the screen (fig. 2). When the oscillator frequency is exactly equal to the frequency of the tuning fork this half circle is stationary, but when there is a difference in frequency it rotates. The best method is to arrange the apparatus so that the tuning knob of the oscillator has to be turned clockwise to increase the frequency and that the half circle rotates clockwise if the oscillator frequency is too high and anti-clockwise if it is too low. Matching is then done very quickly, the vernier control of the oscillator being turned in the direction opposite to that in which the half circle is moving, until the latter is stationary [5]). (A half circle has been chosen here for the sake of simplicity, but this is not essential for matching the frequencies. One can also work with the half of any ellipse, since the phase shift in the network need not be precisely 90°.)

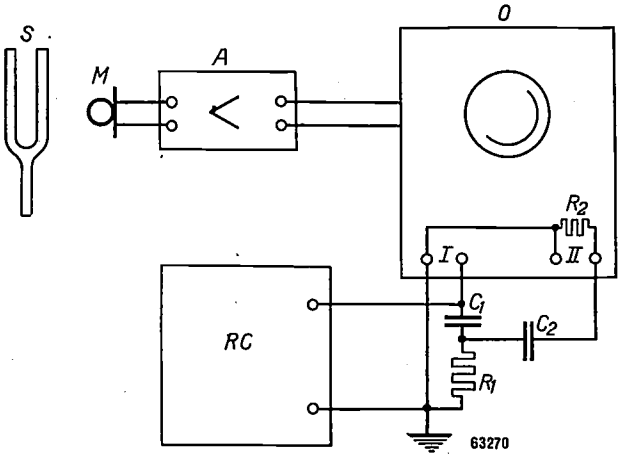


Fig. 2. Circuit for matching the frequency of an *RC* oscillator (*RC*) to that of the tuning fork to be calibrated (*S*). The signal from the oscillator is applied direct to the input terminals *I* of the cathode-ray oscilloscope *O* and, via a 90° phase-shifting network of resistors and capacitors (R_1, C_1, R_2, C_2), to the input terminals *II*. The beam current of the cathode ray tube is modulated by the output from an amplifier (*A*) to which is connected a microphone (*M*) picking up the sound from the tuning fork.

As indicated in the diagram, R_2 may be the resistor already contained in the oscilloscope between the terminals *II*.

[5]) This method has been used before, for measuring the pitch of musical instruments during a musical performance (Balth. van der Pol and C. C. J. Addink, Philips Techn. Rev. 4, 205-210, 1939. and Wireless World 44, 441-442, 1939). The oscillator — in that case a calibrated string with electrically sustained vibration and variable through a small frequency range — could thereby, within one second, be made equal in frequency to the note a^1 in the music to within 0.2 c/s. For calibrating a tuning fork, however, a string variable in frequency is not accurate enough.

If a voltage source with the standard frequency is available (e.g. an electrically driven tuning fork, or the 440 c/s signal as regularly transmitted by the B.B.C. for tuning purposes), then the tuning-fork frequency that is to be determined can be compared directly with the standard frequency by means of the set-up shown in fig. 2, replacing the *RC* oscillator by the standard frequency signal. When this method is followed a rotating half-circle is again seen, and from the number of revolutions it makes in a certain time it is easy to derive the difference in frequency (the sign is indicated by the direction of rotation). Thus one measures, as it were, a fraction of a beat. The accuracy is proportional to the duration of the observation, being already very great at the duration of the decay time. If, for instance, the half circle makes $^1/_4$ revolution in 12.5 seconds then the frequency difference is 0.02 c/s.

Measuring the oscillator frequency with the aid of a clock

A synchronous clock with a seconds hand is then connected to the *RC* oscillator matched to the tuning fork. In our case a clock was used running at the right speed when connected to a signal of 1000 c/s. If it is fed with a frequency f during a period of time T_1 and in that time its reading changes by an amount T_2 then

$$f = \frac{T_2}{T_1} \cdot 1000 \text{ Hz}.$$

T_1 is measured with a good pendulum timepiece checked before and after the measurement with radio time signals, which thus form the basis of the calibration. Since the seconds hand of this clock moves forward 1 second at a time, and in fact the two clock readings cannot be taken simultaneously with sufficient accuracy, a stop-watch is used, this being started when the seconds hand of the pendulum timepiece arrives at a certain dial mark, say 60. From that moment onwards the watch forms, as it were, an extension piece of the pendulum clock giving readings accurate to within 0.2 or 0.1 second according to whether there are 5 or 10 balance movements per second. If the watch is stopped when the continuously rotating seconds hand of the synchronous clock passes a certain dial mark then the moment that this has taken place is found by adding the position of the stopped watch to the time shown by the pendulum timepiece. The three readings (of the pendulum clock, the synchronous clock and the stop-watch) are noted down.

After a sufficient length of time the procedure is repeated and from the readings taken one can find immediately T_1 and T_2 (see the figures given below by way of example).

The error in the movement of the pendulum clock could be made smaller than one second per

24 hours, i.e. less than $1/10^5$, so that no correction need be made for this. Neither is any correction necessary for the movement of the stop-watch, since for the purpose of this test it runs for such a short time that any error can be ignored.

The reading errors at the beginning and at the end of the time T_1 being additive, when a stop-watch is used that can be read accurately to within 0.2 s the duration of T_1 is known accurately to within 0.4 s. Since, as deduced above, the measurement with a tuning fork of about 440 c/s has to be accurate within about 0.01%, the duration of T_1 has to be at least $0.4/10^{-4} = 4000$ s, thus about one hour. (With a stop-watch recording time to within 0.1 s the duration of T_1 need be only half an hour.)

A picture of the complete calibrating apparatus is given in *fig. 3*.

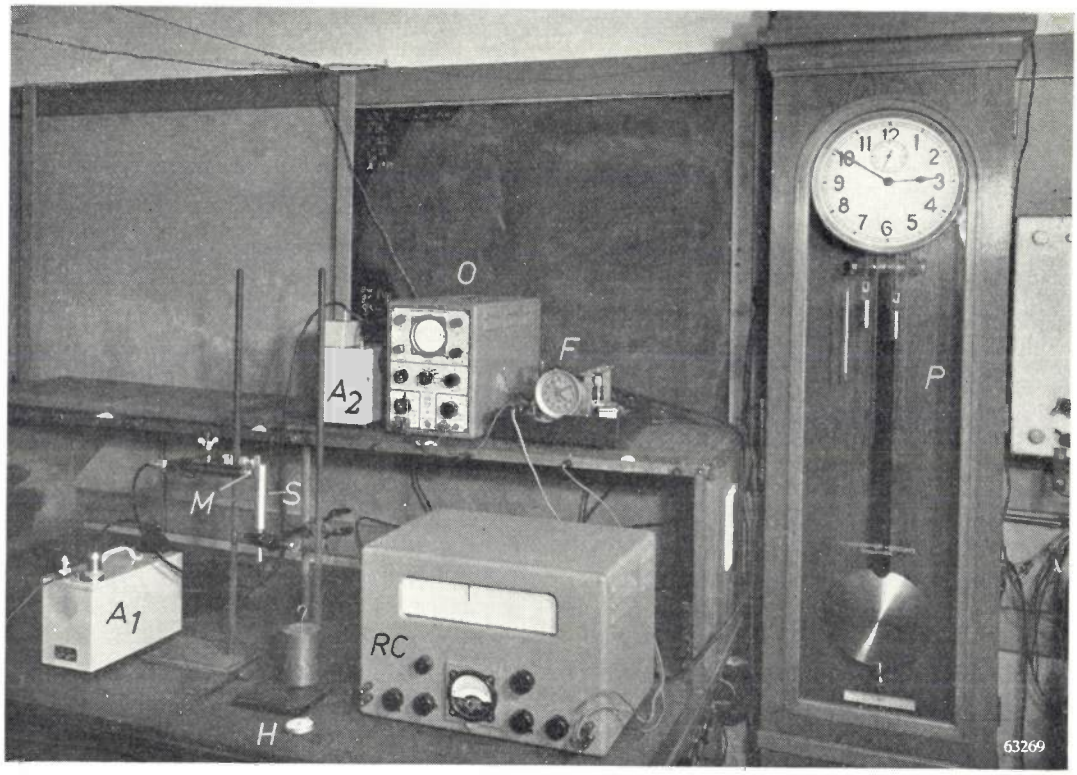


Fig. 3. Complete apparatus for calibrating tuning forks. S the tuning fork to be calibrated, M microphone, A_1 and A_2 amplifiers together forming the amplifier A of fig. 2, O oscilloscope, RC auxiliary oscillator, P pendulum timepiece, F synchronous clock, H stop-watch.

Some readings taken during a practical calibration are given below.

The watch was started when the pendulum clock showed 2 hrs 37′ 0″ and stopped when the seconds hand of the synchronous clock passed a given mark of the dial, viz. when this clock showed 3 hrs 55′ 35″. The stop-watch then showed 7.2″. Thus at the moment that the synchronous clock showed 3 hrs 55′35″ (the beginning of T_2) the time shown by the pendulum clock was 2 hrs 37′7.2″ (the beginning of T_1). After about one hour, at 3 hrs 31′0″ on the pendulum clock, the stop-watch was started again and then stopped when the synchronous clock showed 4 hrs 19′20″ (end of T_2). The stop-watch then showed 7.8″ and was thus stopped at the time 3hrs 31′7.8″ on the pendulum clock (end of T_1).

Thus we find: $T_1 = 3$ hrs 31′7.8″ — 2 hrs 37′7.2″ = 3240.6″, and $T_2 = 4$ hrs 19′20″ — 3 hrs 55′35″ = 1425″. Thus $f = (1425/3240.6) \times 1000$ c/s = 339.73 c/s.

To minimize any inaccuracy due to errors in reading, the start and the finish of T_1 and T_2 are repeated twice; of the three readings thus obtained the identical ones are usually correct.

Correction of the tuning fork

If the tuning fork tested is found to deviate from the desired frequency then one will want to correct it. If the frequency is too low this can be done easily; the exact frequency can be approached very closely by removing a little material from the end of the prongs of the fork. This has to be done in such a way however, that no difference in the lengths of the prongs is introduced, for the slightest difference considerably increases the damping. For this reason, bifurcurated rotationally cylindrical tuning forks were adopted (fig. 1). The prongs can be shortened as required by stopping the saw cut temporarily with, for example, a strip

of brass, and then shaving the end face in a lathe.

Once the frequency f of the tuning fork has been measured, then using the formula giving the frequency as a function of the dimensions and material constants, an accurate calculation of the requisite modification can be made. The formula is [6]):

$$f = \frac{K^2 v}{(l + l_0)^2},$$

where K^2 is a factor dependent upon the shape of the tuning fork, v is the velocity of sound in the material of the fork, l the length of the prongs and l_0 a constant, which in the present case is small compared with l (a few %). The length of the prongs is taken to be the distance from the vibration node to the tip; with the form described above the vibration node lies on one level with the centre of the bore at the bottom of the saw cut.

After the prongs of a fork having too low a frequency have been shortened by the amount calculated according to this formula, the remaining frequency deviation is usually less than the tolerance allowed ($\pm$ 0.2 c/s).

If the tuning fork has a frequency slightly too high it is still possible to correct it, by making the prongs a little thinner at the level of the vibration node. It can hardly be calculated in advance, however, how much material has to be removed, and furthermore the appearance of the tuning fork is thereby spoiled. For this reason the uncorrected tuning forks are always given a frequency slightly too low.

This type of fork is made in a lathe and a milling machine. A rod is turned to produce the required diameters for the prongs and the stem, and symmetry is assured. A hole is now drilled at a right angle through the axis of the rod. A slot, in line with the drilled hole, is cut along the axis of the rod on a milling machine. It is obvious that this is a much simpler process than that of making the usual rectangular cross section fork.

So as to have a light and stainless product a hard aluminium alloy was chosen instead of steel, although it has a somewhat greater absolute temperature coefficient.

A disadvantage due to the shape and the material chosen is that the fork has a weaker tone and a shorter vibration time than the usual fork when placed on a sounding board. However, this fork gives better results when held close to the ear, especially at low frequencies, owing to the relatively larger radiating surface of the prongs.

Summary. A method is described by means of which tuning forks can be calibrated without requiring an acoustic standard frequency. With the aid of a cathode-ray oscilloscope the frequency of an RC oscillator is matched to that of the tuning fork to be calibrated. Connected to this oscillator is a synchronous clock, the seconds hand of which would make one revolution per minute if the frequency were 1000 c/s. If, after an interval T_1, this clock shows a difference in reading T_2 then the frequency sought is $f = (T_2/T_1) \cdot 1000$ c/s. The duration of the interval T_1 is measured with the aid of a pendulum timepiece the movement of which is checked before and after the measurement with radio time signals.

If the frequency found is lower than that required then the latter can be very closely approximated by shortening slightly the prongs of the tuning fork. A special, cylindrically shaped tuning fork is described, with which this correction can be made easily on a lathe, thus retaining the required symmetry.

[6]) See, e.g., Handbuch der Physik, part. 8, p. 207 (Springer, Berlin 1927).

EXPERIMENTAL TESTING OF ELECTRICAL NETWORKS BY MEANS OF THE UNIT FUNCTION RESPONSE

by J. van SLOOTEN. 621.317.75:621.317.333.8

As is known, the response of an impedance or an electrical network to the so-called unit function current or voltage gives valuable information about the nature of the impedance or the network. Hitherto, however, very little use has been made of this for answering questions which repeatedly occur in experimental work in a radio laboratory, such as: How great is the inductance of this coil? What is the damping of that circuit? Is this filter terminated with the right resistance? A simple measuring device, using the method of unit function response, has been developed in order to meet this need.

Unit function response curve

In the theory of electrical networks particular attention has long been paid to the reaction of a network to the sudden switching on or off of a direct voltage or a direct current at the input of the network, i.e. to the response to a voltage or a current in the form of a so-called unit function. The graphical representation of this response as a function of time is called a unit function response curve. It is known that this curve tells us just as much as the frequency response curves describing the amplitude and phase of the response of the network as a function of the frequency [1]). Mathematically speaking, the unit function response curve is therefore equivalent to the frequency response curves. This does not mean to say, however, that in practice it is immaterial which of these curves is available. In the television and radar techniques one frequently has to deal with pulses of finite duration and it has for long been known that in such cases the unit function response curve is more valuable than the frequency response curves: the reaction of a network to a pulse of finite duration is found by taking the difference between two equal unit function response curves, one of which is shifted over the width of the impulse with respect to the other.

Complicated instruments for studying the response of networks to pulses of short duration have already been developed [2]). In very many cases, however, a much simpler apparatus will be able to render just as good service, not only for studying fairly complex networks but also for all kinds of routine measurements. Particularly if no stringent requirements are imposed as regards the accuracy of the measurements — for example in the case of sampling tests in the manufacture of radio parts like coils, I.F. tranformers, etc. — much time can be saved by employing a simple measuring device as described below.

By means of this simple device the response of an impedance or a network to a current in the form of a unit function is traced on the screen of a cathode-ray tube. From the oscillogram thus obtained it is possible to draw various conclusions, as some examples will show. The fact that the equipment can be so simple is due to the use of a unit function instead of pulses. The response to a pulse, as already remarked, can easily be deduced from the oscillogram if such is desired.

When the response of a network to a current impulse (of finite duration) is compared with the response to a unit function of current of the same amplitude, these responses are only of the same order if the duration of the pulse is of the order of one cycle of the response (if this is periodical) or of the time constant of the response (if it is non-recurring). This implies that in an apparatus working with impulses both the interval between the impulses and the duration of the impulses have to be separately variable. In principle, therefore, such an apparatus requires a more complicated system than one working with unit functions of current.

Description of the apparatus

The apparatus consists of a device supplying the unit function of current to be sent through the impedance or network under test, and a cathode-ray tube with its accessories.

Since the response has to be investigated by means of an oscillogram, the unit function has to be repeated periodically, and for this purpose use is made of the well-known Abraham and Bloch multivibrator. This multivibrator can be so

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

[2]) See, e.g., D. S. Espley, E. C. Cherry and M. M. Levy, The pulse-testing of wide-band networks, J. Inst. El. Engrs. **93** III A, 1176-1187, 1946.

designed as to furnish not only a rectangular voltage with the desired frequency — which voltage controls a valve supplying the rectangular current — but also a sawtooth voltage required for the horizontal deflection of the electron beam in the cathode-ray tube.

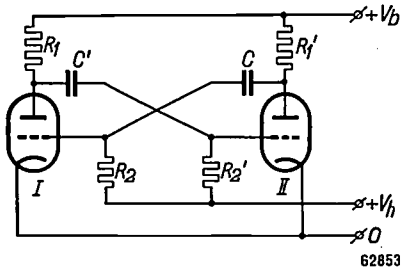


Fig. 1. Basic circuit of the multivibrator. I and II triodes (or pentodes), R_1 and R_1' anode resistors, R_2 and R_2' grid resistors, C' and C coupling capacitors, V_b supply voltage, V_h auxiliary voltage.

The circuit diagram of the multivibrator is shown in *fig. 1*. The two valves I and II are interconnected in such a way that the voltage drop in the anode circuit of one valve keeps the other valve cut off, and vice versa. As soon as some current begins to flow through the valve first cut off (owing to decrease of the negative grid voltage), say I, the positive feedback between the two valves causes the anode current in valve I to increase further. As a consequence the anode voltage of valve I drops, the grid of the valve II becomes more negative and the anode current in valve II decreases until that valve is entirely cut off. This process is then repeated in the reverse direction, and so on. The transition of current from one valve to the other is so rapid that the variations of the anode voltage of the valves are approximately rectangular.

So long as a valve of the multivibrator is conducting its grid voltage is practically zero. While the valve is cut off its grid voltage changes according to an exponential function, but if the grid resistors are connected to a point of high positive voltage then that function is practically linear, because only the first, almost rectilinear, part of the exponential curve is traversed. The positive voltage referred to (V_h in fig. 1) may be equal, for instance, to the supply voltage V_b (300 V) of the anode circuits.

Fig. 2 represents the variations of the voltages on the anode and on the grid of one of the valves. The most important deviation from a rectangular form of the anode voltage is indicated by the dotted curves; at the moment that the valve (say I) is cut off the anode voltage does not immediately become equal to the supply voltage V_b. The reason for this is that in the other valve there is at first a fairly strong grid current, drawn mainly from the coupling capacitor (C' in fig. 1), causing a voltage drop in the anode resistor (R_1). This has the effect of making the shape of the anode voltage somewhat less rectangular at the moment the valve is cut off than at the instant when it becomes conductive. This sharp change in the voltage when the valve becomes conductive is therefore used to bring about a sudden drop of the anode current in a third valve, the switching valve, to zero. The input impedance of the network to be tested is incorporated in the anode circuit of the switching valve (a pentode) and in this way the unit function response (the voltage across the impedance or across the output terminals of the network) can be observed on the screen of the oscilloscope.

In principle there is no necessity for the third valve, because the impedance to be investigated could be directly connected in series with one of the anode resistors of the multivibrator, since unit function currents also occur in these resistors. Unless, however, the impedance is very small compared with the anode resistances, the working of the multivibrator is disturbed. Hence the desirability of using a separate, third valve as switching valve.

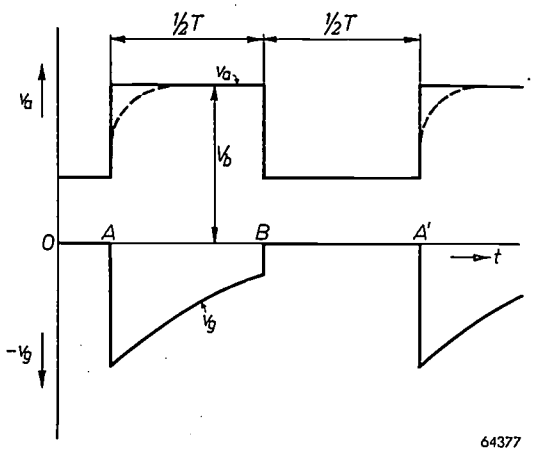


Fig. 2. Anode voltage v_a and grid voltage v_g of one of the multivibrator valves (on different scales). When the valve is cut off (at A) the dotted curve is obtained, which does not occur when the valve begins to pass current (at B).

The frequency of the multivibrator has to be such that the switching-off transient to be investigated almost completely dies out within half a cycle. Since the current flowing through the switching valve and the impedance under investigation changes alternately from zero to a certain value I_{max} and back again to zero, there is alternately a switching-on and a switching-off transient. In order to avoid intermixing of these two phenomena on the screen of the oscilloscope, the switching-on transient is either rendered entirely invisible or

compressed to a vertical line at the side of the oscillogram; this will be reverted to later.

Fig. 3 is a complete circuit diagram of the oscilloscope with its accessories. The impedance to be investigated, X, is connected between the terminals p and r. With the switch S_2 in the position shown in the drawing the voltage across the impedance is applied to the plates for the vertical deflection. When a four-terminal network has to a conducting connection between the points q and r, so as to keep the average potential difference between these points at zero level; the resistance of this connection may be high, up to about 1 megohm.

The frequency of the multivibrator, and therefore the duration $T/2$ of a half cycle in which the oscillogram is described, is varied by switching on (with S_1) coupling capacitors of different capacitances.

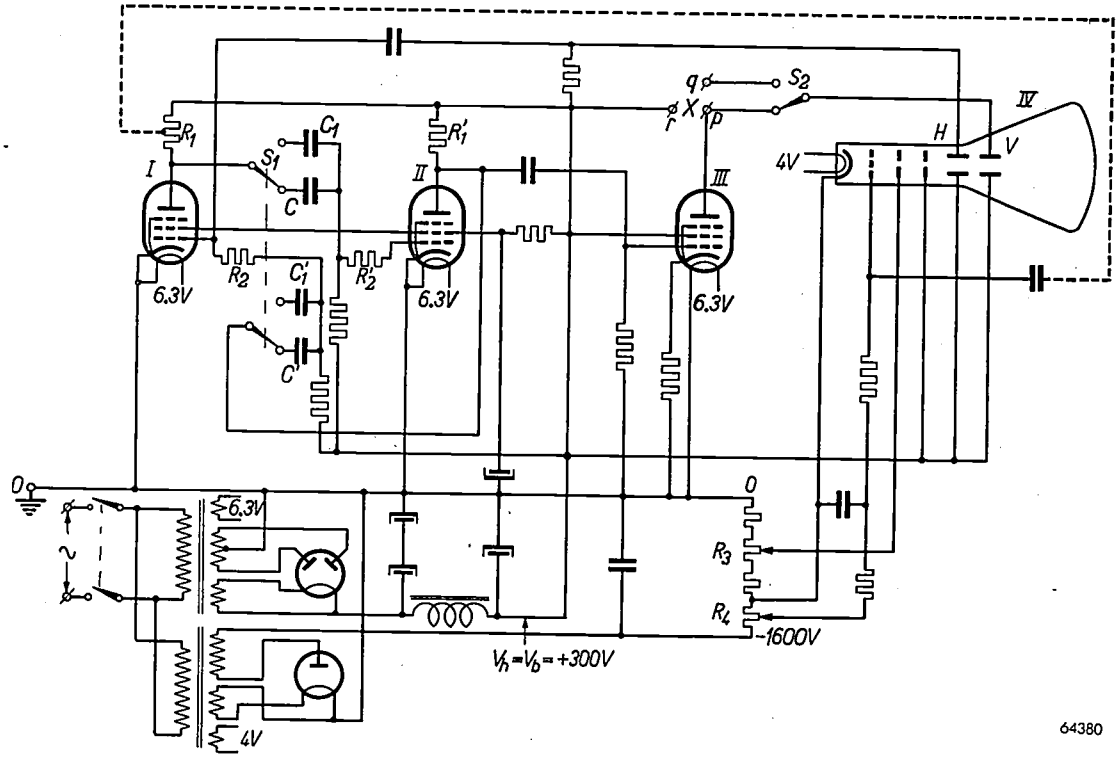


Fig. 3. Complete circuit of the oscilloscope. *I*, *II* multivibrator valves (type EL 41), *III* switching valve (EL 41), *IV* cathode-ray tube (DG 10-3), R_1 and R_1' anode resistors, R_2 and R_2' grid resistors of the multivibrator. Two different frequencies are obtained by connecting the coupling capacitors C and C' or C_1 and C_1' by means of the switch S_1. The terminals p and r are connected to the impedance X under investigation (switch S_2 in the position drawn) or to the input of a four-terminal network under investigation, the output of which is connected to q (S_2 in the upper position). H plates for the horizontal deflection, V plates for the vertical deflection. The electron beam is focused with R_3 and the brightness of the oscillogram is adjusted with R_4.

be investigated then the input terminals are connected to p and r and one of the output terminals to q; with S_2 in the other position the voltage variation at the output terminal is then depicted on the screen.

The direct-current resistance between the terminals p and r must not be higher than about 5000 ohms, in view of the maximum permissible voltage drop (150 V) caused by the D.C. anode current (30 mA); sufficient voltage has to be left for the proper functioning of the switching valve *III*. Further, it is necessary to make sure that when the switch S_2 is in the upper position there is always

Suitable values for $T/2$ are, for example, 1000 μs and 40 μs. It is of importance to know this duration exactly, and the easiest way to determine it is to connect a resistor across the terminals p and r and then to compare on an oscilloscope the frequency $1/T$ of the rectangular voltage across that resistor with the frequency of a calibrated oscillator. Another (less accurate) method, dispensing with an oscillator, will be described later.

One of the plates for the horizontal deflection (see fig. 3) is connected to one of the grids of the multivibrator valves via a coupling capacitor. From fig. 2 it is seen that during one half cycle the grid

voltage rises almost linearly, so that a likewise practically linear time base is obtained, and that during the other half cycle the grid voltage is about zero. The latter means that when the switching-off transient of the impedance under investigation is produced on the screen, the picture of the switching-on transient is compressed to a single vertical line on the extreme right of the picture, where it is not disturbing (in many cases it is even useful, as will be shown presently with some examples). If desired the vertical line can be entirely suppressed by applying to the control grid of the cathode-ray tube, during the respective half cycle, a negative voltage which can be taken from the multivibrator by means of the connection indicated by the dotted line in fig. 3.

Some examples will now be given to show how the oscilloscope described here can be used to advantage.

Examples of application

Resistance

The simplest impedance to be investigated is of course a resistance, one of say 3000 ohms. The oscillogram produced will then have the appearance of that given in *fig. 4*. During the half cycle that the control grid voltage of the valve I (fig. 3) is about zero (BA', fig. 2) the spot is somewhere on a vertical line. At the beginning of this half cycle the switching valve suddenly passes current and the resulting voltage drop in the resistor R to be investigated causes the spot to travel upward [3]). Thus the switching-on transient appears on the screen as a vertical line, the length of which is proportional to R. At the end of that half cycle (at A', fig. 2) two

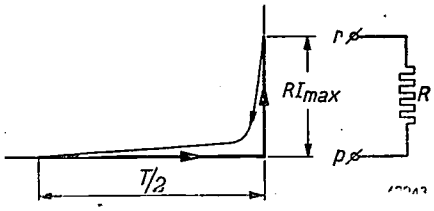


Fig. 4. Oscillogram obtained from a resistor R connected to the terminals p and r (fig. 3). The length of the vertical line on the right is proportional to the resistance R, the anode current amplitude I_{max} and the sensitivity of the cathode-ray tube.

things happen simultaneously: the current through the switching valve is suddenly cut off and the grid voltage of the valve I suddenly assumes a high negative value. As a consequence the spot of light very rapidly moves downward and across to the left, the exact course it follows being immaterial (in fig. 4 this is indicated by a lightly drawn line). In the half cycle then starting, the voltage across the resistor remains zero, while the grid voltage of the valve I gradually increases, the spot thus describing a horizontal line from left to right and returning to the starting point at the moment that the next switching-on transient sends it upward again, and so on [4]).

The "sensitivity" of the oscilloscope can be determined by connecting a known resistance. It is in the order of 1 cm deflection per 1000 ohms. The length of the vertical line obtained with a certain resistance can be compared with the oscillogram of composite impedances.

Resistance and capacitance in parallel

When a capacitor C of comparatively low value is shunted across a resistor R ($RC \ll T/2$) an oscillogram is obtained similar to that shown in *fig. 5*. On the right is again the compressed picture

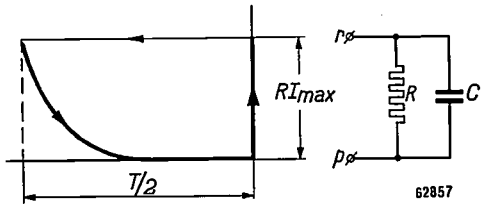


Fig. 5. Oscillogram obtained from a resistor R shunted by a small capacitor C ("small" means $RC \ll T/2$).

of the switching-on transient. When the current in the switching valve has dropped from I_{max} to zero the voltage v across the R-C circuit decreases exponentially with the time t according to the function:

$$v = R\, I_{max}\, e^{-\frac{t}{RC}} \quad \ldots\ldots \quad (1)$$

During the transition of current from one multivibrator valve to the other, the electron beam in the cathode-ray tube first moves rapidly from the top right-hand corner of the oscillogram to the top left-hand corner and then more slowly downward to the bottom right-hand corner. If the oscilloscope has been carefully mounted, so that stray capacitances are kept low, this rapid flyback takes no longer than about 1 μs and thus this interval of time is negligible compared with $T/2$.

[3]) Here and in the following examples it is assumed that the plates for the vertical deflection are connected in such a way that the voltage drop caused by the current in the impedance to be investigated results in an upward deflection.

[4]) With the process described here the oscillogram comes to lie roughly in the middle of the screen, since a blocking capacitor ensures that the mean value of the deflection voltage for the horizontal direction is zero.

If the duration of the stroke $T/2$ is known, it can be estimated how long it takes for the voltage v to drop to 1/e (about 37%) of the peak value. Since this time is equal to RC (see eq. (1)), when the value of R is known a rough estimate can be made of the capacitance C.

Conversely, with known values of R and C the duration of the stroke $T/2$ can be determined, without having to use the calibrated oscillator referred to above. The greatest accuracy is obtained if RC and $T/2$ are of the same order. The oscillogram then assumes the shape represented in *fig. 6*.

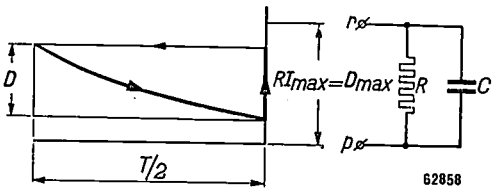


Fig. 6. Oscillogram obtained from a resistor R shunted by a capacitor C in the case where RC is not small with respect to $T/2$.

During the half cycle that the current is switched off the voltage v remains well above zero and while the current is switched on it does not reach the value RI_{max}. A simple calculation shows that the vertical amplitude D on the screen then has the value:

$$D = D_{max}\frac{1-p}{1+p}, \quad \ldots\ldots \quad (2)$$

where $p = e^{-T/2RC}$ and D_{max} is the amplitude corresponding to RI_{max} obtained when C is of a very low value. If C is so chosen that $D = \frac{1}{2}D_{max}$ then from eq. (2) it follows that $p = 1/3$, so that

$$\tfrac{1}{2}\,T = RC\log_e 3 = 1.1\,RC\,.$$

Thus, with R and C known, it is possible to determine the duration of the stroke $T/2$.

Resistance and self-inductance in parallel

Let us now consider the case where an inductance L is connected in parallel with a resistance R and the ratio L/R is small compared with $T/2$.

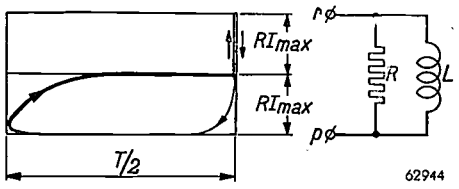


Fig. 7. Oscillogram obtained from a resistor R shunted by an inductor L in the case where $L/R \ll T/2$.

The oscillogram then gives a picture similar to that in *fig. 7*. Owing to the delay caused by the inductance the current continues to flow through L for some time after the current in the switching valve has dropped to zero. This gives rise to a positive voltage across R, and thus, under the conditions assumed in footnote [3]), a downward deflection, which decreases exponentially according to the formula:

$$v = RI_{max}\,e^{-\frac{R}{L}t}.$$

In a manner similar to that described above for capacitances, inductances can thus be roughly measured, the most accurate results being obtained by comparison with one or more known inductances.

L-C circuit

The oscillogram obtained with an L-C circuit is that of a damped oscillation (*fig. 8*), provided the damping of the circuit is not so great as to cause the phenomenon to be aperiodic.

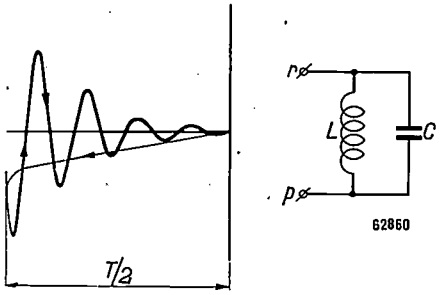


Fig. 8. Oscillogram obtained from a damped oscillatory circuit.

The formula for the voltage v across the capacitor is:

$$v = V_0\,e^{-\frac{r}{2L}t}\cos\omega_0 t,$$

where V_0 is the initial value of v, r is the loss resistance taken to be concentrated in the inductance, and $\omega_0 = 2\pi$ times the natural frequency of the circuit.

The damping of the circuit can easily be determined from the ratio of the successive peak values of v. Inductances can also be measured by comparing the number of oscillations in a certain interval of time with the number occurring with the same capacitance and a known inductance.

Band-pass filters (I.F. transformers)

The band-pass filters so frequently employed in radio technique can also be studied with the aid of the unit function response curve. A band-pass filter may consist of two L-C circuits coupled, for

instance, by a mutual inductance M (see *fig. 9*, where for the sake of simplicity the two circuits are assumed to be identical).

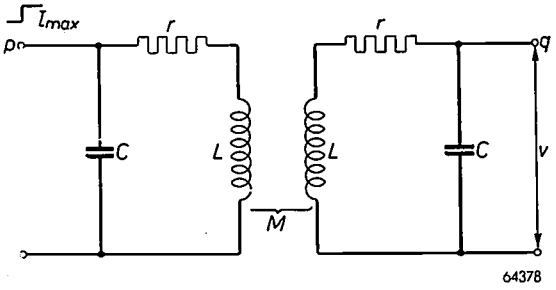


Fig. 9. Band-pass filter (I.F. transformer) with two identical circuits coupled by a mutual inductance M.

A characteristic of a band-pass filter is its coupling factor k, which , for the filter represented in fig. 9, is defined as M/L. When the behaviour of the filter is investigated as a function of k one finds a remarkable difference between the method of the unit function response and that of the frequency response. When the latter method is followed, thus measuring the amplitude of the output voltage as a function of the frequency, it is found, as is known, that with a loose coupling the frequency response curve shows one, rather low, peak at the angular frequency $\omega_0 = 1/\sqrt{LC}$, whilst as the coupling becomes gradually tighter the peak rises until at a certain critical coupling it cannot rise any higher and splits up into two peaks lying either side of ω_0. The value k_{cr} of the critical coupling factor is equal to the damping factor δ:

$$k_{cr} = \delta \equiv r/\omega_0 L = r\sqrt{\frac{C}{L}}.$$

For the variation (as a function of time) of the secondary voltage v of a band-pass filter, when a unit function of current I_{max} is applied to the primary terminals, the following approximate expression can be formulated:

$$v = I_{max}\sqrt{\frac{L}{C}} \cdot e^{-\frac{r}{2L}t} \cdot \sin \tfrac{1}{2}k(\omega_0 t - \delta) \cdot \cos(\omega_0 t - \delta). \quad \ldots \quad (3)$$

(It is approximate because both the damping factor δ and the coupling factor k are taken to be small, e.g. $\delta < 0.1$ and $k < 0.1$, which is usually the case.)

From the eq. (3) it is seen that the secondary circuit oscillates at its own angular frequency ω_0 (the cosine term) and that this oscillation is modulated with the much lower angular frequency $\frac{1}{2}k\omega_0$ (the sine term); this is also to be seen from the graphical representation in *fig. 10*. Further, the formula shows that the amplitude of the oscillation undergoes the natural damping (term with the e-power).

The term $\sin \frac{1}{2}k(\omega_0 t - \delta)$ corresponds to the energy flowing to and fro between the two circuits. In fig. 10 it is seen that the amplitude reaches a number of maxima. The ratio of two successive maxima depends upon the value of δ/k. The time τ between two maxima is given by $\frac{1}{2}k\omega_0\tau = \pi$, and hence:

$$\tau = \frac{2\pi}{\omega_0 k}.$$

The ratio a of a maximum in the envelope drawn in dotted lines (fig. 10) to the preceding maximum is therefore:

$$a = e^{-\frac{r}{2L}\cdot\frac{2\pi}{\omega_0 k}} = e^{-\frac{\pi\delta}{k}}. \quad \ldots \quad (4)$$

Whereas from the shape of the frequency response curve of a band-pass filter it can immediately be seen whether the coupling is looser or tighter than the critical coupling (one peak or two peaks in the curve), the unit function response curve (fig. 10) does not provide any such simple criterion, since when passing the critical coupling this curve does not noticeably change in shape.

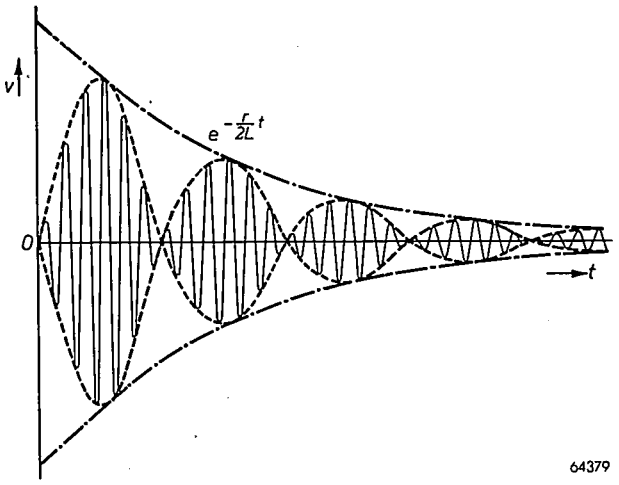


Fig. 10. Variation of the secondary voltage v of the band-pass filter in fig. 9 when a unit function current is applied to the primary.

This, however, does not prevent conclusions about the coupling being drawn from the oscillogram of the unit function response curve. According to eq. (4) the ratio a increases uniformly as the coupling becomes tighter and at the critical coupling assumes the value $e^{-\pi} \approx 1/23$, that is to say, the second and following maxima are then already so small as to be almost indistinguishable in the oscillogram. If, however, the oscillogram shows marked and different maxima, as in fig. 10, then

it is to be concluded that the coupling is very much tighter than the critical coupling. By measuring the ratio a from the oscillogram and applying formula (4) one can then determine with reasonable accuracy the ratio δ/k.

Low-pass filter

As a last example, we shall take the case of a low-pass filter, assuming that the output is terminated by a resistance R and that the input is open-circuited (*fig. 11*). The oscillograms obtained of the output voltage when the terminating resistance is too high, correct or too low are of the types represented in figures 11*a*, *b* and *c* respectively. If a resistance is applied also across the input of the filter and this resistance is approximately equal to the characteristic impedance of the filter, similar figures will be obtained but the oscillations are then much weaker. The most favourable value of the resistance is that with which the oscillations are as weak as possible, and with the aid of the oscilloscope this value can easily be determined experimentally.

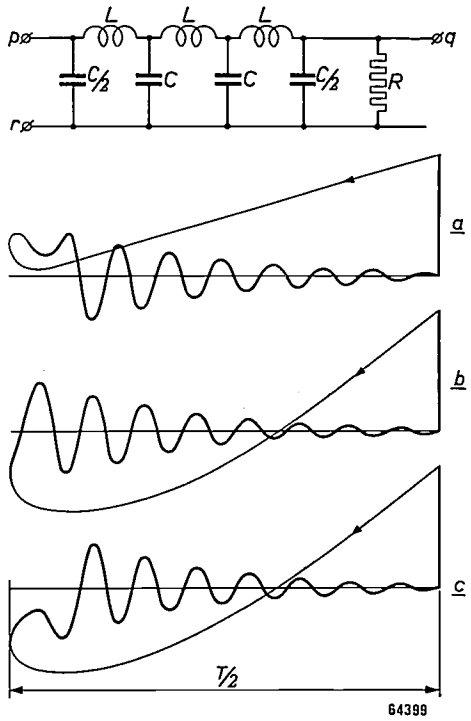


Fig. 11. *Top:* a low-pass filter terminated by a resistor R. When a unit function of current is applied to the input the output voltage varies according to *a*, *b* or *c*, depending upon whether R is too high, of the correct value, or too low respectively.

In the same way it is possible to investigate the effect of unequal filter sections, for which otherwise complicated calculations would be required.

Although these investigations by means of an oscilloscope and the unit function response method do not yield any very accurate quantitative results, this method has the advantage that it reveals very quickly which circuit elements in a particular case are of the most importance and which are less essential. Its application is particularly advantageous in small laboratories where the means available are restricted (both as regards equipment and in respect to personnel).

Summary. A simple experimental method is described for studying the response of an electrical network or that of an impedance to the sudden switching off of a constant direct current. The voltage produced by the switching off is observed as a function of time on the screen of a cathode-ray oscilloscope. The switching off is done periodically with the aid of a multivibrator, which supplies at the same time the time-base voltage for the oscilloscope. The voltage variation observed permits of a number of conclusions being drawn in regard to the network or the impedance, as several examples show. This method is particularly suitable in cases where an answer is required quickly and not too stringent quantitative requirements have to be met.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherland.

1932: G. H. Jonker and J. H. van Santen: Ferromagnetic compounds of manganese with perovskite structure (Physica **16**, 337-349, 1950, No. 3).

Various "manganites" of the general formula $La^{3+}Mn^{3+}O_3^{2-}$, $Me^{2+}Mn^{4+}O_3^{2-}$ (Me representing a bivalent metal) have been prepared in the form of polycrystalline products. Perovskite structures were found, i.a., for all mixed crystals $LaMnO_3$-$CaMnO_3$, for $LaMnO_3$-$SrMnO_3$ containing up to 70% $SrMnO_3$, and for $LaMnO_3$-$BaMnO_3$ containing less than 50% $BaMnO_3$. The mixed crystals with perovskite structure are ferromagnetic. Curves for the Curie temperature versus composition and saturation versus composition are given for $LaMnO_3$-$CaMnO_3$, $LaMnO_3$-$SrMnO_3$, and $LaMnO_3$-$BaMnO_3$. Both types of curves show maxima between 25% and 40% $Me^{2+}Mn^{4+}O_3^{2-}$; here all third electrons available contribute with their spins to the saturation magnetization. The ferromagnetic properties can be understood as the result of a strong positive Mn^{3+}-Mn^{4+} exchange interaction combined with a weak Mn^{3+}-Mn^{3+} interaction and a negative Mn^{4+}-Mn^{4+} interaction. The Mn^{3+}-Mn^{4+} interaction, presumably of the indirect exchange type, is thought to be the first clear example of positive exchange interaction in oxidic substances.

1933: R. Loosjes, H. J. Vink and C. G. J. Jansen: The potential distribution in pulsed oxide-coated cathodes and its consequences for the velocity distribution of the emitted electrons (J. Appl. Phys. **21**, 350-351, 1950, No. 4).

Previous investigations (see Nos **R 124** and **1910** of these abstracts) have shown that a potential difference of e.g. 200 V may exist across the oxide layer of an oxide-coated cathode at a current density of 10 A/cm² and that this potential drop is concentrated near the outer surface of the layer. In consequence of this the velocities of the emitted electrons show much greater spread than would be expected if the distribution were Maxwellian. The velocity spectrum of the electrically deflected electrons shows two (sometimes three) "lines" instead of being continuous, the spread amounting to 230 eV with $V_a = 860$ V, $I_a = 14$ A/cm², $T = 942$ °C.

R 137: G. Diemer and J. L. H. Jonker: On the time delay of secondary emission (Philips Res. Rep. **5**, 161-172, 1950, No. 3).

The frequency limit of a disc-seal dynatron proves to be 2400 Mc/s. From this an upper limit of 3×10^{-11} sec for the time lag of secondary emission is derived. Measurements on the admittance of an U.H.F. dynatron at 300 Mc/sec show that the time lag must be smaller than 10^{-11} sec. From transit time effects within the secondary-emitting material an upper limit for the time lag of the order of 10^{-14} to 10^{-15} sec may be estimated theoretically.

R 138: E. J. W. Verwey, P. W. Haaijman, F. C. Romeijn and G. W. van Oosterhout: Controlled-valency semiconductors (Philips Res. Rep. **5**, 173-187, 1950, No. 3).

Elaborate treatment of the subject already referred to under No. **1895**.

R 139: R. A. Hutner, E. S. Rittner and F. K. du Pré: Fermi levels in semiconductors (Philips Res. Rep. **5**, 188-204, 1950, No. 3).

General formulae for determining the Fermi level and the density of free charge carriers in semiconductors are derived. Special semiconductor models are considered in detail and a few applications are discussed.

R 140: J. M. L. Janssen: A cathode-ray oscillograph for periodic phenomena of high frequencies (Philips Res. Rep. **5**, 205-240, 1950, No. 3).

An oscillograph is described for periodic signals of high frequency, working according to a scanning principle. By mixing the high-frequency signal with phase-modulated pulses a faithful A.F. image is obtained. The bandwidth and the maximum scanning speed of the oscillograph are investigated as functions of the pulse frequency and pulse width. Further, the synchronization circuit, the mixer circuit and the pulse generator are described (c.f. Philips Techn. Review **12**, 52-59 and 73-82, 1950, No. 2/3).

VOL. 12 No. 9, pp. 241-272 MARCH 1951

Philips Technical Review

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

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

THE SYNCHROCYCLOTRON AT AMSTERDAM

I. GENERAL DESCRIPTION OF THE INSTALLATION
II. THE OSCILLATOR AND THE MODULATOR

by F. A. HEYN*).

621.384.61

Reports on cyclotrons often have rather the appearance of aiming at overwhelming the public with amazement, with all their talk about metres thick magnet poles, hundreds of tons of iron, tens of kilowatts, velocities approaching the velocity of light, etc. Here, in this description of the Philips synchrocyclotron, it will be shown — without any attempt at belittling the impressiveness of the installation — how it comes about that in the building of a cyclotron one is in many respects obliged to adopt exceptional methods of construction.

I. GENERAL DESCRIPTION OF THE INSTALLATION

Philips have built a synchrocyclotron which has been installed in the Institute for Nuclear Research at Amsterdam, where it was officially taken into use on 10th November 1949. This apparatus is now working day after day and only a few interruptions have been necessary for maintenance. It is operated by Philips' personnel and for two-thirds of the time is at the disposal of Philips for their own research work and for the production of artificial radioactive substances used in all sorts of medical, biological and technical investigations. For the rest of the working time the beam of acccelerated particles produced by this apparatus is available for scientific research in general. With the approval of the Institute also research workers at universities in the Netherlands can avail themselves of the opportunities offered by the synchrocyclotron for carrying out their investigations.

In this first article (I) a concise general description of the construction of this synchrocyclotron will be given, together with a brief account of what has so far been achieved with it. Further articles will deal with the most important component parts of the installation, the first of these (II) following immediately upon the general survey.

Tbe theoretical principles of the cyclotron were discussed in an article (further referred to as A) which appeared in this journal some months ago [1]). For the benefit of the reader the most fundamental facts and formulae will be repeated here where necessary.

We shall start by giving in *figs 1a* and *1b* a schematic drawing of the chief part of the installation, the actual accelerating apparatus. *Fig. 2*, a photograph of a model of the apparatus in a partly dismantled state, will make it easier to understand the drawing.

The drawing shows the evacuated accelerating chamber *1*, in which the particles, in our case either deuterons (ions of heavy hydrogen) or alpha particles (doubly ionized helium atoms), emitted by the

*) Professor at the "Technische Hogeschool"(Engineering University) at Delft (Netherlands), formerly of Philips Laboratories at Eindhoven.

[1]) W. de Groot, Cyclotron and synchrocyclotron, Philips Techn. Rev. 12, 65-72, 1950 (No. 3).

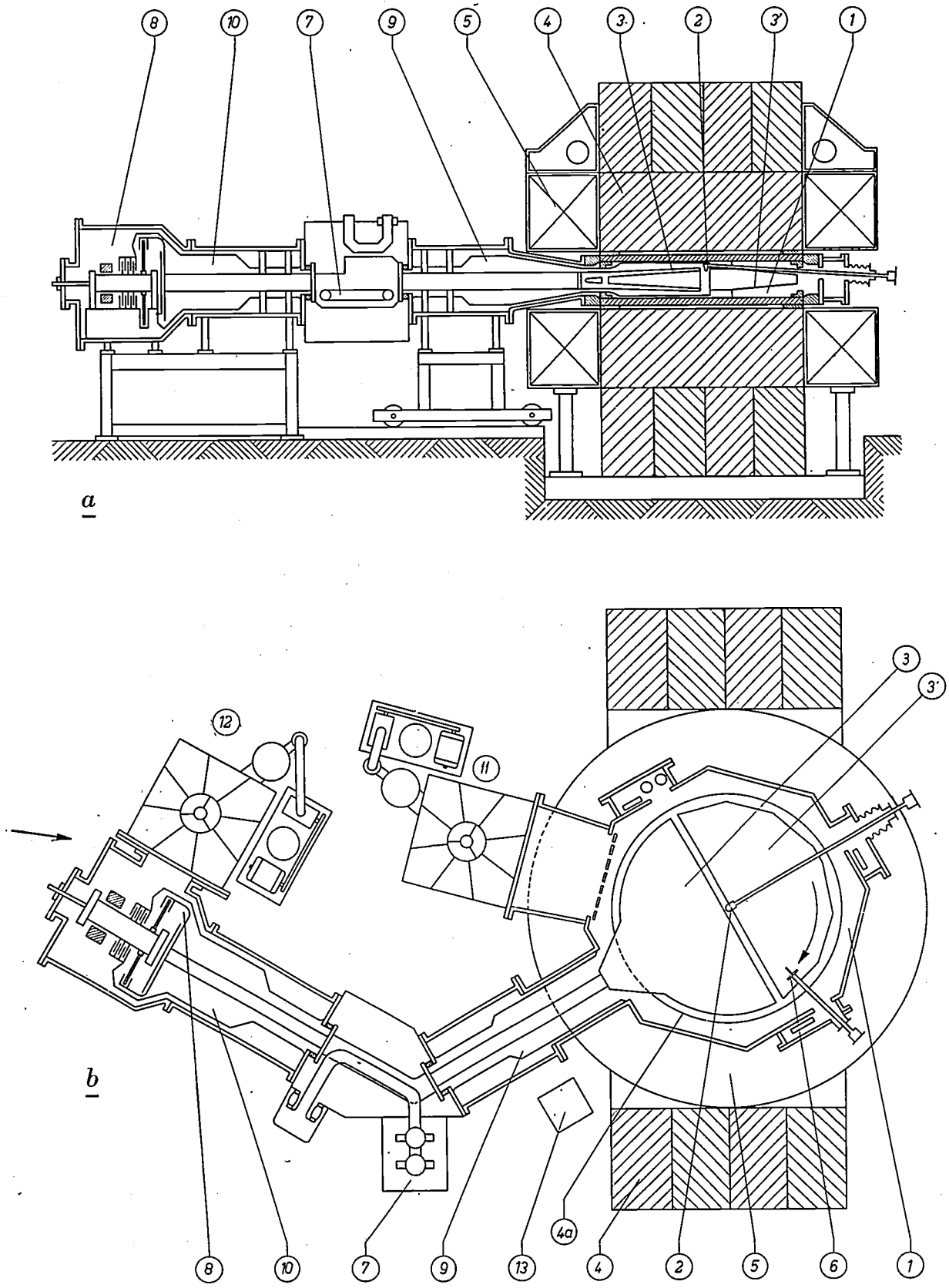


Fig. 1. Greatly simplified vertical and horizontal cross sections of the Philips synchrocyclotron at Amsterdam. (The vertical cross section, *a*, is really a juxtaposition of two cross sections along the centre lines of *9* and *10* in *b*.) *1* evacuated acceleration chamber, *2* ion source, *3* and *3'* dees, *4* magnet, *4a* pole face, *5* energizing coils, *6* target, *7* valves and other parts of the oscillator, *8* modulator, *9* and *10* coaxial transmission line, *11* and *12* vacuum pumps, *13* boron counter chamber.

ion source *2* are to be accelerated, by means of an alternating voltage of the order of 15 kV_{peak} between the dees (*3* and *3'*). The magnetic field of the large magnet *4*, energized with the aid of the coils *5*, causes the particles to describe a circular path, at such a rotational speed that they return each time in the right phase to the gap between the dees to undergo again an accelerating action from the alternating voltage, but with the

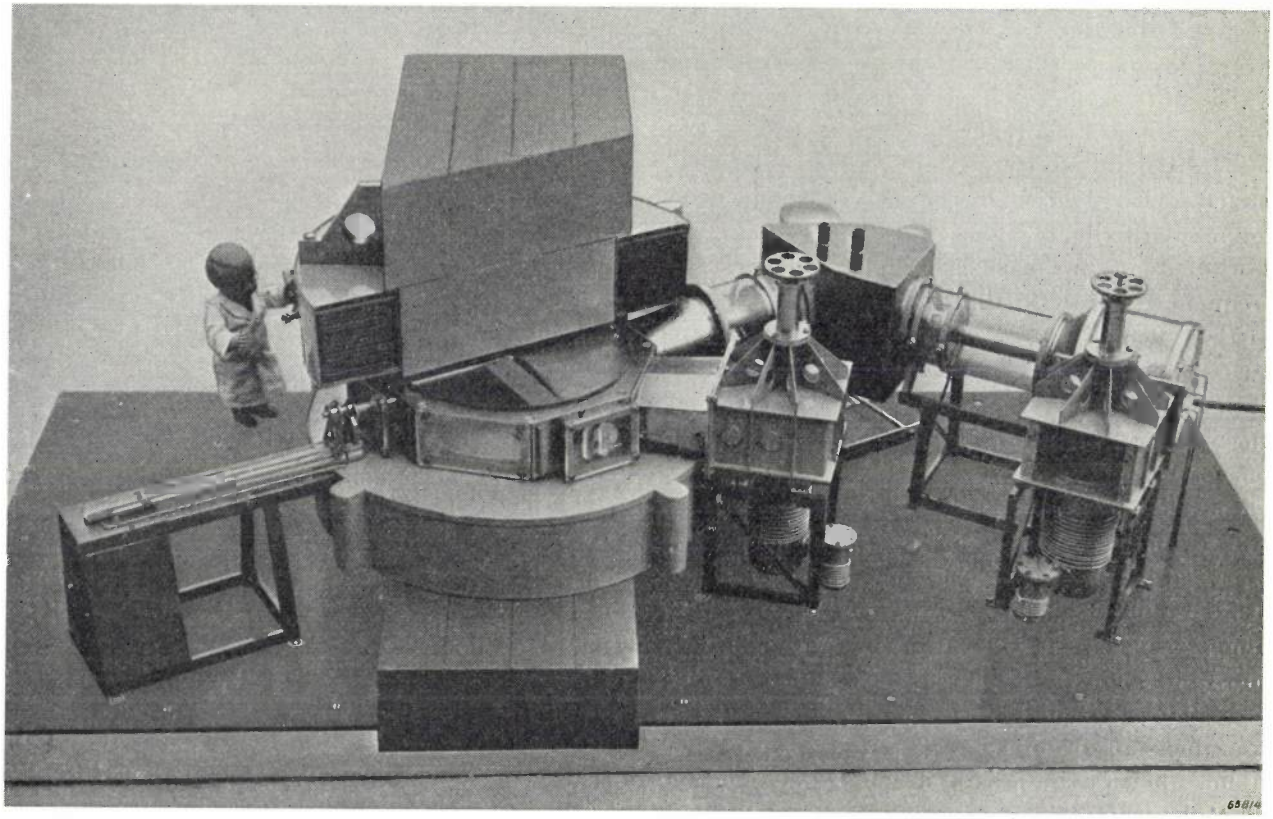


Fig. 2. View of the synchrocyclotron, with the magnet partly cut away to show the acceleration chamber with the two dees.

radius of the orbit increasing with the increasing energy of the particles. The ultimate energy to be reached is determined entirely by the flux density B of the magnetic field and the ultimate radius r of the orbit [2]). In our case $B = 1.38$ Wb/m^2 (13,800 gauss) and $r = 78$ cm, so that deuterons can be accelerated to a final energy of 28 million electron volts and alpha particles to 56 MeV.

To generate the powerful magnetic field mentioned above in the large gap between the pole pieces (diameter of the pole pieces 1.80 m, distance between pole pieces 36 cm) 200 tons of iron was required for the magnet, whilst the energizing coils contain 32 tons of copper and are fed with 180 amperes. The electric power required to compensate the I^2R-losses is 80 kW and oil cooling is provided to dissipate this power. The great distance between the pole pieces (great also in comparison with similar installations) was chosen so as to be able to produce a very intensive beam of particles. This was particularly important for the object aimed at in the designing of this synchrocyclotron. The relation between pole distance and beam intensity is made clear when bearing in mind that, although by a suitable field variation the ideal path of the particles is stabilized in the centre plane of the gap, the particles may still oscillate vertically (and radially) about that ideal path. The more space is left for the vertical oscillations, i.e. the higher the accelerating chamber, the more particles can escape collision with the walls of the dees and thus participate in the acceleration right to the end.

The beam of accelerated particles is made to strike the target *6*, which can be inserted more or less deep into the accelerating chamber

[2]) See, e.g., Philips Techn. Rev. **11**, 70, 1949 (No. 3).

to adjust the energy of the particles collected. The materials to be transmuted can be put on the target. Due in part also to a suitable design of the ion source, it has already been possible to increase the beam intensity of our synchrocyclotron so that the average intensity of current carried by the charged particles (deuterons) to the target amounts to 20 micro-amperes. This is to be regarded as a high value. With this current on the target an energy of $28\times10^{6}\cdot20\times10^{-6} = 560$ watts is converted into heat, which has to be dissipated by water cooling (it is by measuring this heat that we get to know the current strength). Also the dees and many other parts are cooled with water.

The alternating voltage required between the dees is obtained with the aid of an oscillator (7 in fig. 1). For the acceleration of particles with mass m and a charge q the alternating voltage must have an angular frequency of $\omega_E = qB/m$ (see article A). For deuterons as well as for alpha particles the frequency required, with the aforementioned value of B, is therefore 10.7 Mc/s. For accelerating protons twice this frequency would be needed.

Since it seemed impracticable at the time to make the oscillator suitable for both these frequencies, one of the first steps in the designing of our synchrocyclotron was to decide with what particles it was to be worked. Fortunately this choice is not very critical: both with deuterons and with protons a large number of nuclear reactions can be brought about which yield a great variety of radioactive substances, among which are materials that cannot be produced from a uranium pile. Weighing the usefulness of protons against that of deuterons, it may be said that with an energy of some tens of MeV deuterons are certainly more suitable for producing neutrons and that in general they may perhaps offer more universal possibilities of application. Anyhow there was no doubt that if the (lower) deuteron frequency were chosen there would be fewer difficulties to overcome in the construction of the oscillator and, particularly, of the modulator, so that in our case it was decided in favour of the deuterons.

The modulator is indicated in fig. 1 by *8*. Let us recall briefly the function of this important part (a detailed explanation is given in A). As soon as the particles reach a high energy in the accelerating chamber they have a tendency to "get out of step": firstly because they get closer to the edge of the pole pieces, where the flux density B must necessarily be smaller than in the middle in order to stabilize the path of the particles, and secondly because the relativistic increase of their mass then becomes noticeable. Owing to these two causes the condition for synchronism $\omega_E = qB/m$ is then no longer fulfilled, the angular frequency ω_E being too high for the particles at the edge. This limits the final energy attainable with the classical cyclotron. When, however, ω_E is gradually reduced — the principle of the synchrocyclotron — the particles can be accelerated to greater final energies. The desired frequency variation is brought about with the modulator. This apparatus contains a capacitor the value of which varies periodically with time and acts on the tuning of the oscillator. In that part of each modulating cycle in which the oscillator frequency decreases, a stream of particles is accelerated. An important feature is the phase stabilization that takes place, in consequence of which it is not necessary that the reduced frequency should at every instant be exactly in step with the changing situation as the radius of the orbit of the individual particles increases. It is due to this that with the synchrocyclotron powerful intermittent beams of accelerated particles can be obtained.

The modulation frequency in our apparatus is 2000 c/s, which means to say that 2000 times per second a group of accelerated particles reaches the target. This relatively high modulation frequency, which is favourable for a high average intensity of the beam current — the number of accelerated particles per group being in the first instance independent of the modulating frequency —, involved no great technical difficulties on account of the fact that in our case the frequency sweep need only be 4%. In fact, from the middle to the edge the flux density drops in our apparatus by about 2.5%, and the mass increase of a deuteron having received a kinetic energy of about 30 MeV at the edge is about 1.6% (the rest energy of the deuteron is 1840 MeV), so that with a frequency drop of about 4% the condition for synchronism $\omega_E = qB/m$ is again fulfilled at the edge [3]).

For the sake of comparison some figures are given relating to the synchrocyclotron at Harwell in England, which was put into service some time ago [4]). This accelerates protons to an energy of about 180 MeV. The relativistic mass increase is then already about 20% and the frequency sweep has to be about 30%. The modulation frequency has for the time being not been made greater than 80 c/s.

[3]) This is likewise the case for alpha particles, since the rest energy and the maximum kinetic energy obtained are in this case both twice these energies of the deuteron, so that the mass increase is again about 1.6%.

[4]) T. G. Pickavance, J. B. Adams and M. Snowdon, The Harwell cyclotron, Nature **165**, 90-91, 21st January 1950.

It is well to emphasize once more (cf. A) that the higher final energy attainable is not the only advantage of the synchrocyclotron compared with the classical cyclotron. In our case the final energy of the deuterons is 28 MeV, whilst with a classical cyclotron deuterons can also be accelerated up to 25 MeV with a not unreasonably small efficiency. The synchrocyclotron, however, requires a much lower oscillator power for such a final energy and also in other respects the construction is much simpler than that of a classical cyclotron, as will be seen in the next article in this series.

ted on an undercarriage running on rails laid radially to the magnet. After the part *7* of the oscillator has been removed parts *9* and *3* can be run out without having to shift part *10* of the transmission line and the modulator.

Further it is to be seen in fig. 1 that there are two vacuum pumps with their backing pumps, one (*11*) connected to the accelerating chamber, the other (*12*) to the modulator. The reason why one pump alone was not sufficient will be seen later.

Finally the synchrocyclotron installation comprises all sorts of parts, some of them very bulky,

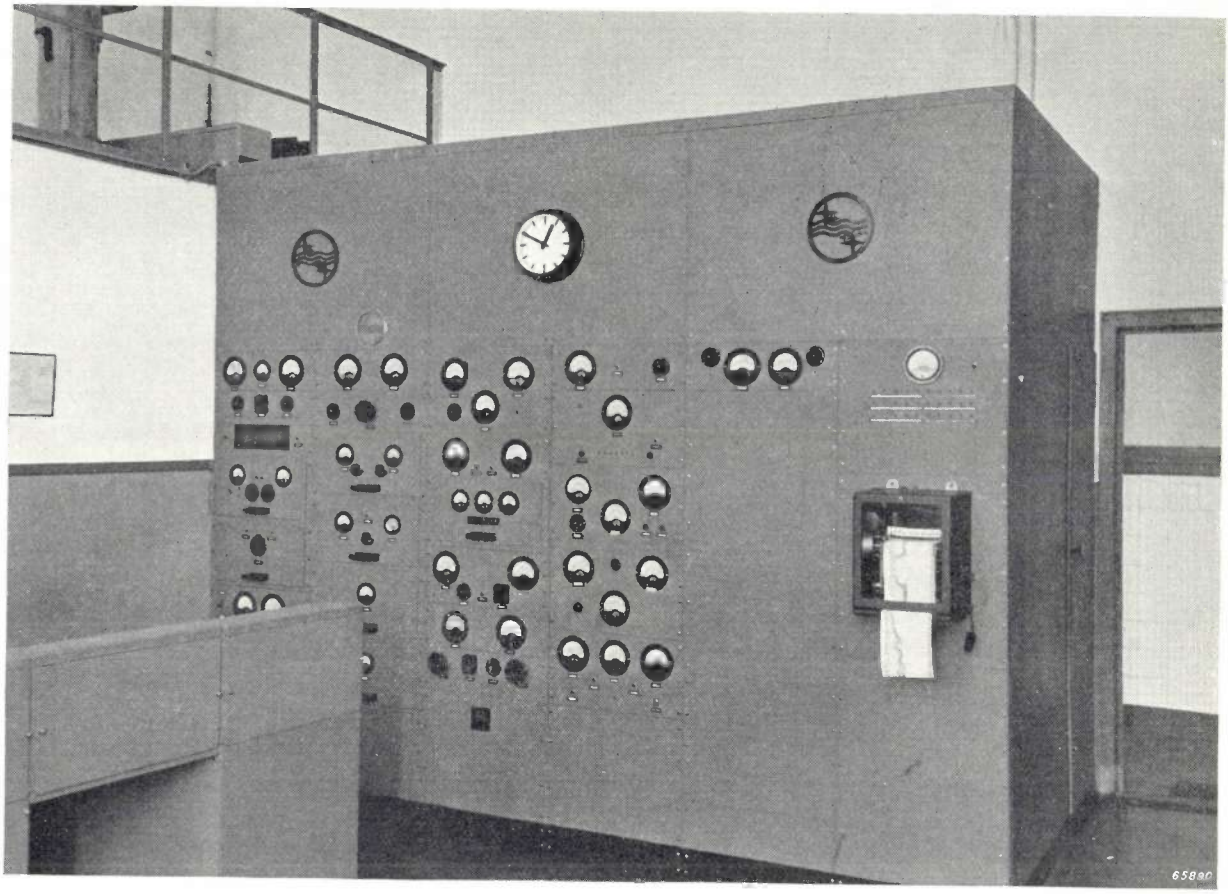


Fig. 3. Part of the control and operating room of the synchrocyclotron at Amsterdam. The personnel in this room are protected against the radiation from the accelerating apparatus by a layer of water 3.5 m thick.

Let us turn back to fig. 1 again. The high-frequency oscillation is applied to the dees and the modulator via a coaxial transmission line, *9* and *10*; the fact is that at the frequency used (wavelength about 30 m) and with the large dimensions of the parts — to begin with the accelerating chamber —, the system can no longer be regarded as a lumped circuit. The whole system is set up with an angle in the middle instead of in one straight line; the only reason for this is that, for the purpose of maintenance, part *9* of the transmission line with the dee *3* should be easily removable. To that end the part *9* is mounted on an undercarriage running on rails laid radially to the magnet. After the part *7* of the oscillator has been removed parts *9* and *3* can be run out without having to shift part *10* of the transmission line and the modulator.

Further it is to be seen in fig. 1 that there are two vacuum pumps with their backing pumps, one (*11*) connected to the accelerating chamber, the other (*12*) to the modulator. The reason why one pump alone was not sufficient will be seen later.

Finally the synchrocyclotron installation comprises all sorts of parts, some of them very bulky, corresponding to what is needed for a radio transmitter, these parts not being visible in figs 1 and 2. For instance, there is a rectifying installation supplying the direct voltage of 12 kV for the oscillator valves; a drum with 2 × 10 metres of rubber hose for taking up the potential difference of the cooling water, coming from the water mains, and the cooled anode at a high positive potential; a control and operating room, with a large number of controls and measuring instruments (*fig. 3*). This room is separated from the cyclotron by a water tank 3.5 m thick protecting the personnel against the neutron

and gamma rays generated in the accelerating chamber (in the target and the walls). On the other sides the cyclotron is enclosed by concrete walls 1 m thick. When entering the cyclotron compartment, after the apparatus has been switched off, to collect the products of the transmutations by withdrawing the target from the accelerating chamber, the necessary precautions are of course also taken. The target is fixed to a holder 1.25 m long. Each holder, after having been used for an irradiation, is first

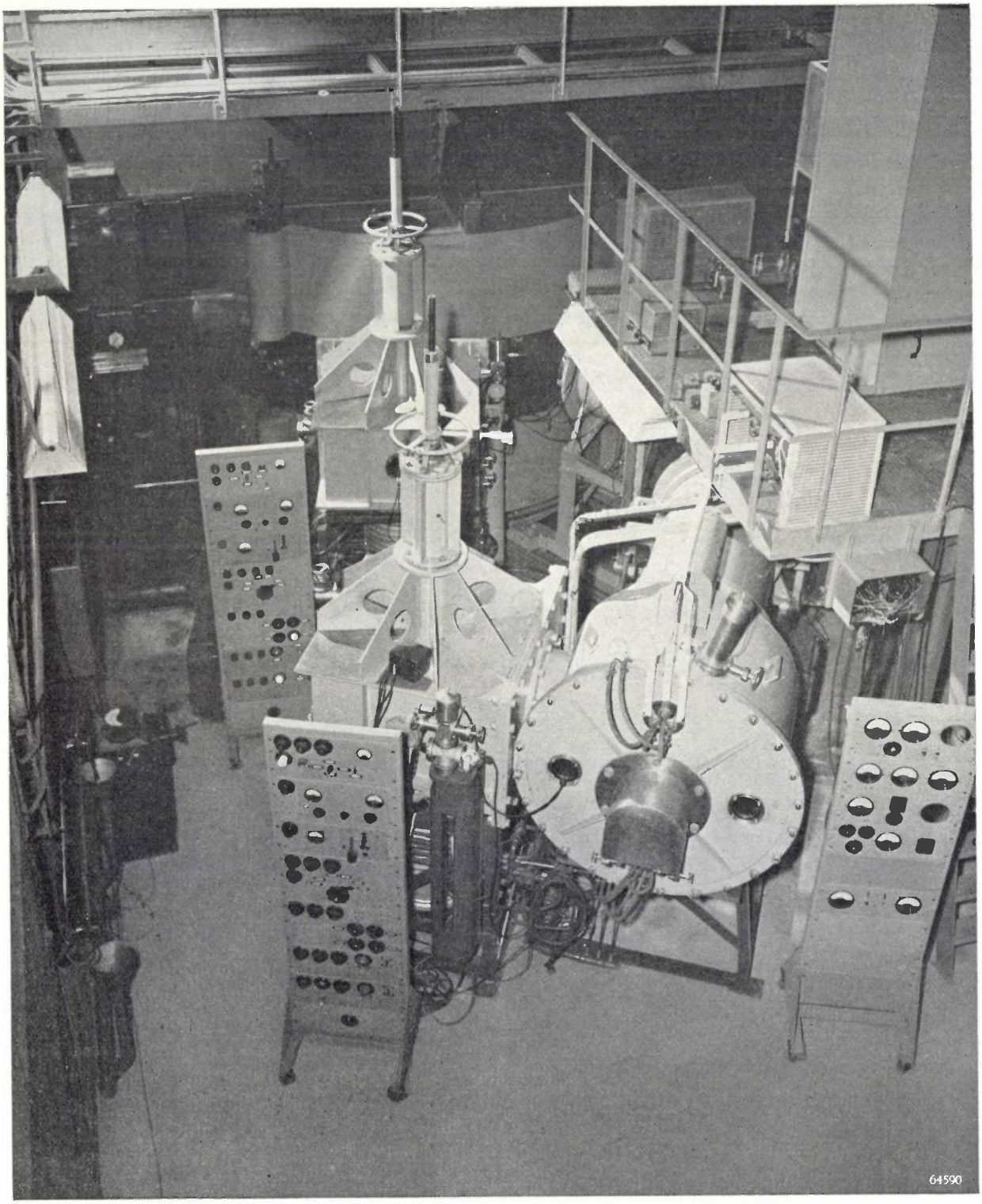


Fig. 4. The synchrocyclotron viewed in the direction of the arrow in fig. 1*b*. In the foreground on the right is the modulator.

left to "cool down" between blocks of concrete for a certain length of time (depending on the "life" and the concentration of the radioactive isotypes formed) before being used again.

The materials are usually irradiated until a radioactivity of some tens of millicuries [5]) has been generated in them. The neutron radiation emanating from the target (and other parts) is being continuously measured with a boron counter chamber [6]) mounted in a fixed position close to the accelerating chamber (*13* in fig. 1).

The photograph in *fig. 4* finally gives a view of the cyclotron as seen in the direction of the arrow in fig. 1*b*. This and the photograph in fig. 3 also give some idea of the extensive accessory apparatus that is needed. In the foreground on the right of fig. 4 is the modulator, which will be discussed in detail in the article now following.

[5]) With phosphorus, for instance, an activity of 150 microcuries per μAh is obtained, thus in this case our beam of 20 μA yields 3 millicuries per hour. The unavoidable losses of radioactive phosphorus in the chemical processing required before the product can be supplied to the "customer" are already discounted in this figure for the yield. (The loss of activity during the chemical processing is negligible considering that the half-value time of this atom is about 14 days.)

[6]) This contains a solid or gaseous boron compound from which alpha particles are released by neutrons. The average ionization current produced by the alpha particles is measured.

II. THE OSCILLATOR AND THE MODULATOR

Principles underlying the construction of the oscillator

Owing to the phase stabilization of the particles travelling round in the synchrocyclotron these can be allowed to describe a large number of revolutions without risk of too many being lost. Consequently for a certain final energy a small gain per "loop" suffices. This is the reason why the alternating voltage between the dees in a synchrocyclotron of the dimensions chosen by us need not be greater than, say, 15 kV_{peak}, whereas for a classical cyclotron of the same dimensions a voltage of certainly 100 kV_{peak} would be required between the dees.

Much smaller voltages than 15 kV_{peak} are also possible, but then there is the objection that the permissible modulating frequency and thus the strength of the beam current likewise become much smaller; this will be reverted to later.

This so much lower voltage has two very important advantages. In the first place there is much less risk of a breakdown of the insulators and disruptive discharges through the gas in the accelerating chamber, and this is an advantage that can be turned to good account in various ways. A higher gas pressure can be permitted in the accelerating chamber, thus making it possible to get greater ion concentrations and therefore more powerful beam currents. Further, one of the two dees can be earthed (thus the full dee voltage, instead of half of it, comes to lie between the other dee and the earthed chamber walls), thereby making it much easier to mount the ion source, the target, etc.

The second advantage of the lower dee voltage is that a much smaller high-frequency power is required than is the case with the classical cyclotron. If the dee circuit has a quality factor Q and a capacitance C (it is imagined for a moment as being replaced by a lumped L-C circuit), then a power

$$P = \frac{\omega C V^2}{2Q} \quad \ldots\ldots\ldots \quad (1)$$

is needed to maintain, in the circuit, an oscillation with peak voltage V and angular frequency ω. (The index E used with ω in article A will henceforth be omitted.) The capacitance C is determined mainly by constructional requirements; in our case it could not be kept below 400 pF. Careful construction allowed of the quality factor Q being raised to 1500. With $\omega = 2\pi \times 10.7 \times 10^6$ and $V = 14$ kV we therefore get $P \approx 1.6$ kW for the dee circuit. The modulator circuit takes up an equal amount of power, so that the high-frequency power required for our synchrocyclotron totals somewhat over 3 kW, as compared with about 100 kW for a classical cyclotron of the same dimensions.

In *fig. 1* a very schematic representation is given of the oscillatory circuit. It may be regarded as a transmission line with two, large, lumped capacitances at the ends: the (non-earthed) dee in the accelerating chamber at one end, with the capacitance $C_1 = 400$ pF, and at the other end an equally large capacitance C_2 (the modulator M). With the aid of an oscillator the transmission line is caused to oscillate in such a way that a standing

left to "cool down" between blocks of concrete for a certain length of time (depending on the "life" and the concentration of the radioactive isotypes formed) before being used again.

The materials are usually irradiated until a radioactivity of some tens of millicuries [5]) has been generated in them. The neutron radiation emanating from the target (and other parts) is being continuously measured with a boron counter chamber [6]) mounted in a fixed position close to the accelerating chamber (*13* in fig. 1).

The photograph in *fig. 4* finally gives a view of the cyclotron as seen in the direction of the arrow in fig. 1*b*. This and the photograph in fig. 3 also give some idea of the extensive accessory apparatus that is needed. In the foreground on the right of fig. 4 is the modulator, which will be discussed in detail in the article now following.

[5]) With phosphorus, for instance, an activity of 150 microcuries per μAh is obtained, thus in this case our beam of 20 μA yields 3 millicuries per hour. The unavoidable losses of radioactive phosphorus in the chemical processing required before the product can be supplied to the "customer" are already discounted in this figure for the yield. (The loss of activity during the chemical processing is negligible considering that the half-value time of this atom is about 14 days.)

[6]) This contains a solid or gaseous boron compound from which alpha particles are released by neutrons. The average ionization current produced by the alpha particles is measured.

II. THE OSCILLATOR AND THE MODULATOR

Principles underlying the construction of the oscillator

Owing to the phase stabilization of the particles travelling round in the synchrocyclotron these can be allowed to describe a large number of revolutions without risk of too many being lost. Consequently for a certain final energy a small gain per "loop" suffices. This is the reason why the alternating voltage between the dees in a synchrocyclotron of the dimensions chosen by us need not be greater than, say, 15 kV_{peak}, whereas for a classical cyclotron of the same dimensions a voltage of certainly 100 kV_{peak} would be required between the dees.

Much smaller voltages than 15 kV_{peak} are also possible, but then there is the objection that the permissible modulating frequency and thus the strength of the beam current likewise become much smaller; this will be reverted to later.

This so much lower voltage has two very important advantages. In the first place there is much less risk of a breakdown of the insulators and disruptive discharges through the gas in the accelerating chamber, and this is an advantage that can be turned to good account in various ways. A higher gas pressure can be permitted in the accelerating chamber, thus making it possible to get greater ion concentrations and therefore more powerful beam currents. Further, one of the two dees can be earthed (thus the full dee voltage, instead of half of it, comes to lie between the other dee and the earthed chamber walls), thereby making it much easier to mount the ion source, the target, etc.

The second advantage of the lower dee voltage is that a much smaller high-frequency power is required than is the case with the classical cyclotron. If the dee circuit has a quality factor Q and a capacitance C (it is imagined for a moment as being replaced by a lumped L-C circuit), then a power

$$P = \frac{\omega C V^2}{2Q} \quad \ldots\ldots\ldots \quad (1)$$

is needed to maintain, in the circuit, an oscillation with peak voltage V and angular frequency ω. (The index E used with ω in article A will henceforth be omitted.) The capacitance C is determined mainly by constructional requirements; in our case it could not be kept below 400 pF. Careful construction allowed of the quality factor Q being raised to 1500. With $\omega = 2\pi \times 10.7 \times 10^6$ and $V = 14$ kV we therefore get $P \approx 1.6$ kW for the dee circuit. The modulator circuit takes up an equal amount of power, so that the high-frequency power required for our synchrocyclotron totals somewhat over 3 kW, as compared with about 100 kW for a classical cyclotron of the same dimensions.

In *fig. 1* a very schematic representation is given of the oscillatory circuit. It may be regarded as a transmission line with two, large, lumped capacitances at the ends: the (non-earthed) dee in the accelerating chamber at one end, with the capacitance $C_1 = 400$ pF, and at the other end an equally large capacitance C_2 (the modulator M). With the aid of an oscillator the transmission line is caused to oscillate in such a way that a standing

wave occurs with at each end a voltage anti-node and in the middle a voltage node (see the sketch). Thus the system forms a half-wave line. Owing to the influence of the lumped capacitances, however, the transmission line need not be 15 m long, which

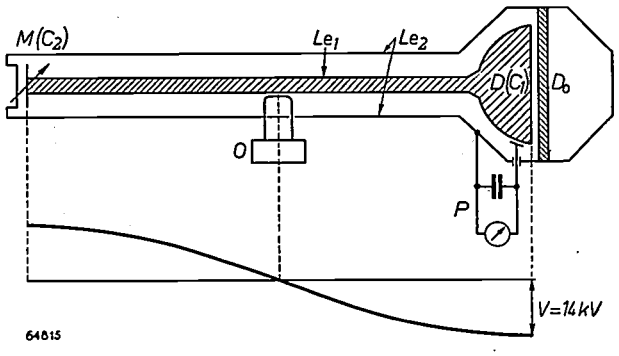


Fig. 1. Very schematic representation of the high-frequency oscillatory circuit of the synchrocyclotron. Le_1, Le_2 coaxial transmission line, D dee, M modulator, O oscillator valve with coupling loop. The other dee (D_0) being earthed, it can in principle be regarded as a narrow strip connected to the walls of the accelerating chamber and placed opposite the edge of the dee under high tension. Where in future we speak of "the dee" this is to be taken as meaning the non-earthed dee. P is a device for measuring the dee voltage. Below this drawing a curve has been plotted showing the voltage distribution along the system.

would correspond to the wavelength of 30 m, but only about 7 m. By periodically varying the capacitance C_2 of the modulator, the frequency is made to fluctuate periodically between about 10.7 and 10.3 Mc/s (variation of 4%, see I).

In order to minimize radiation losses the transmission line has been made in the form of two coaxial conductors, the inner one with a diameter of about 15 cm (of course this inner conductor itself is made in the form of a hollow pipe, since the high-frequency currents flow only along the surface), the other one with a (average) diameter of 60 cm. The peculiar configuration of the transmission line made it impossible to compute precisely in advance the dimensions required for the desired frequency of 10.7 Mc/s. The dimensions of the transmission line and the modulator were therefore determined experimentally with a model (the dimensions of the dee are governed by other considerations, viz. the need to avoid disruptive discharges and to provide as much space as possible for vertical deviations of the travelling particles), for which purpose a full-scale rough model was made which could easily be altered and with which the resonant frequency was determined after every modification. The final form of the model served as pattern for the actual construction, which of course differed from the model in its precise finishing: all parts are made of copper, perfectly fitting and soldered with silver, so that the dissipative resistance of the whole line is no more than 0.05 ohm. In this way the quality factor 1500 was reached.

The necessity of keeping the dissipative resistance so small is evident from a calculation of the current strengths occurring in the system: to charge the 400 pF capacitance C_1 (the non-earthed dee) to a peak voltage of $V = 14{,}000$ volts at a frequency $\omega = 2\pi \cdot 10^7$ a current is needed with the peak value

$$I = \omega C_1 V,$$

i.e. around 350 amperes. With the small total dissipative resistance mentioned, we thus arrive at the previously mentioned total dissipated power of somewhat more than 3 kW. Since the stream of particles impinging on the target generates about 560 watts (it may be possible to increase this figure still further by improving upon the ion source), we may say that about 15% of the high-frequency power is utilized, which for a cyclotron is a remarkably high "efficiency".

The circuit diagram of the oscillator is given in *fig. 2*. A Meissner circuit has been used, i.e. the anode circuit of the transmitting valve B is inductively coupled to the oscillatory circuit (drawn here as a lumped L-C circuit for the sake of simplicity), whilst also the reaction to the grid circuit (excitation) is inductive. The anode and grid

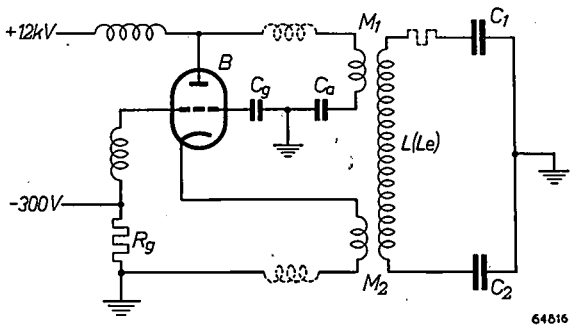


Fig. 2. Circuit of the oscillator. The oscillatory circuit of the synchrocyclotron can be regarded as an L-C circuit with the inductance mainly concentrated in the transmission line (Le) and the capacitance mainly in the dee (C_1) and in the modulator (C_2). This circuit is inductively coupled to the anode and grid circuits of the transmitting valve B (Meissner circuit). Undesired but unavoidable impedances are shown in broken lines.

coupling coils each consist of only one loop, placed beside the inner conductor of the transmission line in such a way that the magnetic lines of force of the loop envelop the conductor, and vice versa.

The Meissner circuit had to be chosen because it is not possible to connect the inner conductor of the transmission line at the point of coupling to a part that has to have a defined (high-frequency) voltage. It is true that at that point, in the middle of the transmission line, there is a voltage node (cf fig. 1), but, owing to the periodical variation of the capacitance C_2 of the modulator, the position of the node is not fixed but shifted a little to and fro along the line. With the inductive couplings according to fig. 2 no trouble is experienced from this.

In order to avoid random oscillations with frequencies determined not by the transmission line but by other L-C circuits occurring in the system, it is necessary that there should be no direct coupling between the anode and the grid circuits. Therefore the inductive coupling between the anode and grid coupling loops was made as weak as possible (see the next section) by giving these loops a suitable shape. Further, also capacitive coupling via the valve is prevented as far as possible by adequately earthing the grid of the valve, for high frequencies, with the large capacitance C_g (in which practically all self-inductance is avoided). Thus the excitation voltage is applied to the cathode, and it would be better to speak of a "cathode coupling loop". The grid bias required is taken from a supply unit with safety device; as soon as the oscillations deviate from the right frequency the grid bias is increased to such an extent that the valve is cut off.

The oscillatory condition of the circuit reads, to a first approximation:

$$(M_1 - M_2)\, M_2 \geqq \frac{r}{\omega^2 S}, \quad \ldots \ldots \quad (2)$$

where M_1 and M_2 are the coefficients of mutual inductance between the coupling loops and the transmission line (fig. 2), r is the circuit resistance already mentioned, and S is the mutual conductance of the valve. Since ω is fixed and r and S cannot be chosen arbitrarily low and high respectively, it was necessary to give $(M_1 - M_2)M_2$ the highest possible value. Thus M_1 has to be as large as possible and $M_2 = \frac{1}{2}M_1$, which is actually the case. Since, however, M_1 and M_2 could only approximately be calculated in advance, it was again necessary to resort to the experimental method with a full-scale model to see whether $(M_1 - M_2)M_2$ would be large enough. By employing two transmitting valves of the type TA 12/20 connected in parallel, it was found possible to get a mutual conductance high enough to fulfil amply the oscillatory condition.

Set-up of the oscillator

Fig. 3 gives a rather more complete and more realistic but geometrically not yet quite exact representation of the oscillator. The two parallel-connected, water-cooled, transmitting valves (here only one is drawn) are mounted in an earthed metal casing divided, by a horizontal partition at grid

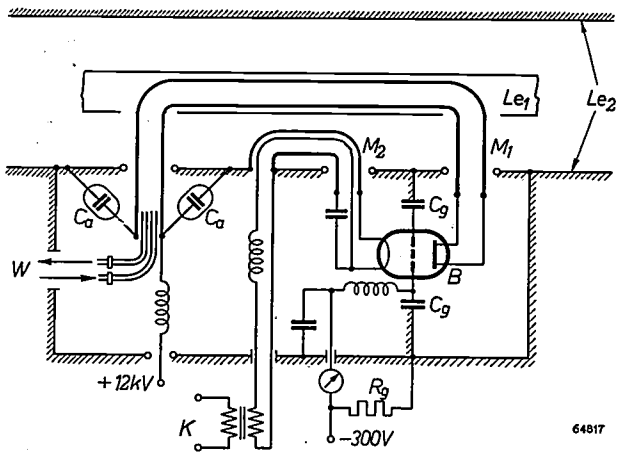


Fig. 3. Detailed circuit of the oscillator with the anode coupling loop M_1 and the cathode coupling loop M_2. For B two transmitting valves, type TA 12/20, connected in parallel are used. C_g capacitors used for by-passing the grid of each valve to earth, C_a capacitors for by-passing the anode coupling loop to earth, R_g grid resistor, K supply for the cathodes, W cooling water for the anodes.

level, into two entirely separated parts, the anode space and the cathode space. The anode coupling loop, a hollow copper conductor with two right-angled bends, passes out of the casing at the side (in the photograph in *fig. 4* it can just be seen) and then runs partly inside a channel of the inner conductor of the transmission line. In this way the desired tight coupling M_1 is obtained, since practically all the magnetic lines of force enveloping one conductor now envelop in this part of the loop also the other conductor, while at the same time the lines of force are kept as short as possible. At its free end the anode loop is earthed for high frequencies by means of a circle of vacuum capacitors (C_a in figs 2 and 3) which introduce hardly any self-inductance and have a very small loss angle, whilst being sufficiently proof against breakdowns to withstand the anode voltage of 12 kV; they can be seen in fig. 4 on the left. The cooling water for the anode is passed in and returned at this point through insulating hoses in the hollow conductor of the loop.

The cathode coupling loop, a copper pipe somewhat shorter and thinner than the anode coupling loop, is led out of the earthed casing and approaches

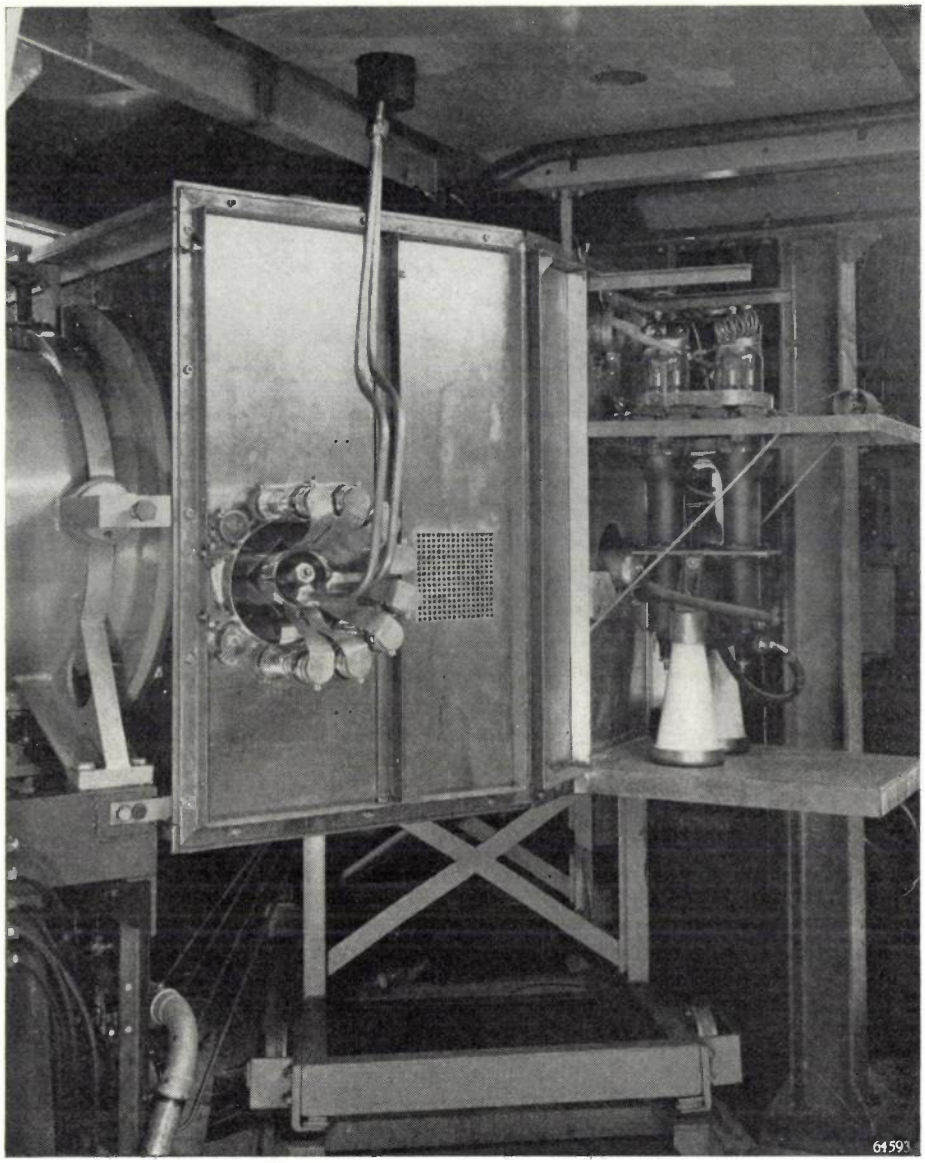


Fig. 4. The central part of the oscillator, opened. The two transmitting valves are seen with the anodes downward. The two hoses for the supply and return of the cooling water pass through the hollow anode coupling loop led through the wall on the left of the valves. Farther to the left is the end of the anode coupling loop connected to the earthed casing by a circle of vacuum capacitors. The connection is made with wide strips so as to keep the inductance low. Also the connection to the water supply pipes (disappearing at the top) is clearly seen. In the lower half of the photograph the undercarriage can be seen that supports half of the transmission line with the dee and runs on rails, so that these parts can easily be run out for any servicing work.

the inner conductor from the top. This could not be drawn in fig. 3, but it is illustrated in the sketch *fig. 5* and can partly be seen in the photograph of *fig. 6*. In this way the coupling between anode and cathode coupling loops is kept small. With the same object in view, the legs of the loops, which contribute towards the mutual coupling but not to the coupling with the transmission line, have been made as short as possible. The useful part of the cathode loop is about half the length of the anode loop, in accordance with the requirement $M_2 = \frac{1}{2}M_1$. The free end of the cathode loop, returning upward, is earthed, and at this point the filament voltage for the cathode is carried through the hollow conductor of the loop.

The grid of each transmitting valve is by-passed to earth by a circle of capacitors (C_g in figs. 2 and 3); here simple mica capacitors could be used, because the requirements as to the breakdown voltage are not so severe.

Although, as stated, the oscillatory condition was amply fulfilled, in the beginning difficulties

were still encountered in the starting of the oscillator. The cause was found to lie in the oscillatory circuit being overloaded by discharge currents flowing in the accelerating chamber from the dee to earth and by electrons reaching the dee from the ion source. By applying a negative direct voltage of some hundreds of volts to the dee, via a choke and the inner conductor of the transmission line (the choke may be seen in the middle of fig. 6), the extra load was sufficiently reduced for the system to be easily brought into oscillation.

It was then found possible to generate an alternating voltage up to 25 kV_{peak} on the dee without any trouble from disruptive discharges in the gas, overheating of parts or overloading of the valves. Already with 5 kV_{peak}, however, the apparatus worked satisfactorily, yielding a reasonable beam of deuterons of 28 MeV. With the normal working voltage of 14 kV_{peak} on the dee the measured input power to the two valves was 5.7 kW and the anode dissipation 2.2 kW, so that the high-frequency output was 3.5 kW and the efficiency of the valves 61%. The value of the quality factor Q calculated from this output and the given values of V, ω and C (see equation (1)) agrees well with the value determined from the width of the resonance curve of the oscillatory circuit. The voltage on the dee is measured by means of a diode voltmeter and a capacitive voltage divider (P in fig. 1).

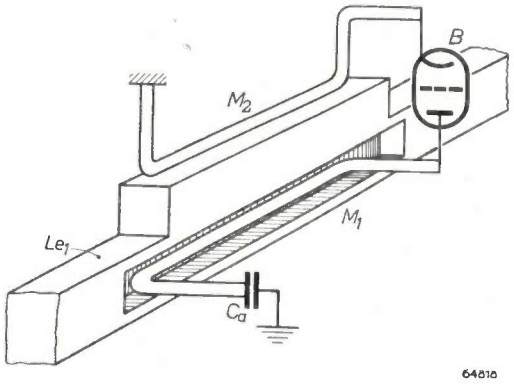


Fig. 5. Configuration of the anode coupling loop (M_1) and the cathode coupling loop (M_2). The inner conductor Le_1 of the coaxial transmission line has a channel in which the anode coupling loop is laid and an elevation where the cathode coupling loop is brought down close to the line. With this construction the mutual coupling between the two loops is kept as small as possible.

The modulator

It has already been mentioned that the required 4% periodical variation of the oscillator frequency is brought about by a variation of the capacitance C_2 (fig. 1). Since approximately all the capacitance

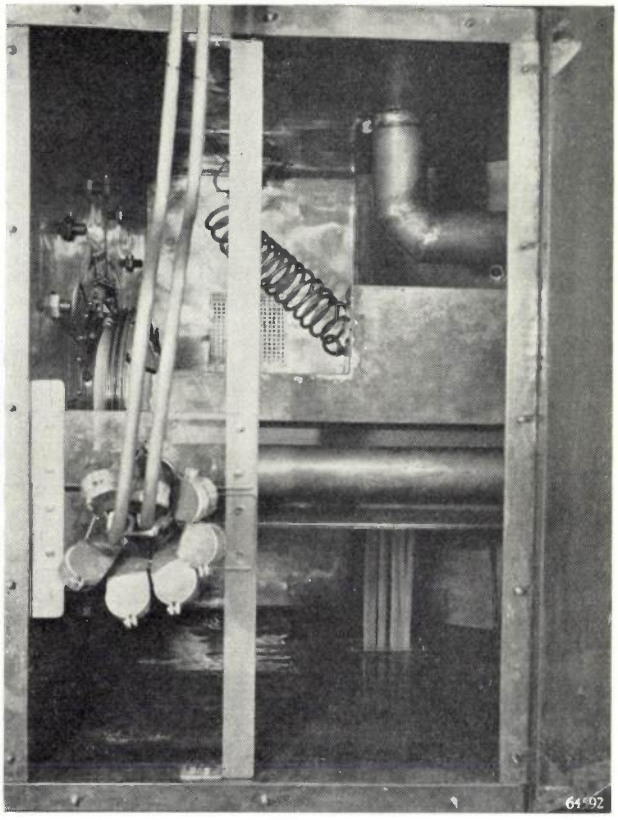


Fig. 6. View inside the middle part of the coaxial transmission line, showing part of the anode coupling loop (at the bottom) and part of the cathode coupling loop (top right). At the bottom on the left is the end of the anode coupling loop with connections for the circle of earthing capacitors and with water pipes.

of the oscillatory circuit is lumped in C_1 and C_2, the total circuit capacitance is given by:

$$C = \frac{C_1 C_2}{C_1 + C_2}.$$

Since in our case $C_2 = C_1$ it follows that

$$\frac{\Delta C}{C} = \tfrac{1}{2}\frac{\Delta C_2}{C_2},$$

whilst further

$$\frac{\Delta \omega}{\omega} = -\tfrac{1}{2}\frac{\Delta C}{C},$$

so we find for the required relative variation of C_2:

$$\frac{\Delta C_2}{C_2} = -4\frac{\Delta \omega}{\omega}.$$

Therefore, to reduce ω by 4%, we have to increase C_2 by 16%, i.e., in the case where $C_2 = 400$ pF, by 64 pF.

This capacitance variation is effected with the aid of a capacitor consisting of a toothed disc rotating between two sets of fixed vanes; see *fig. 7*. When the teeth of the disc are exactly opposite the vanes the capacitance is maximum, and when they are opposite the spaces between the vanes the capacitance is minimum. The disc has 40 teeth and makes 50 revolutions per second, so that the modulating frequency of the dee voltage is 2000 c/s.

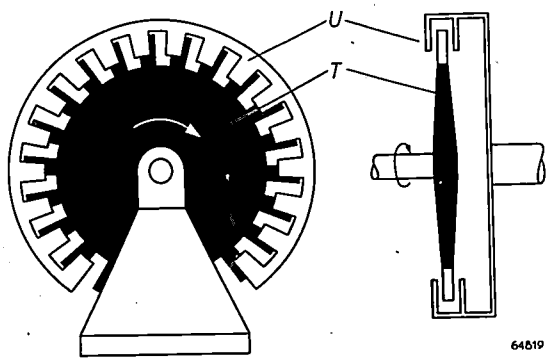


Fig. 7. The modulator comprises a periodically varying capacitance obtained with a toothed disc T rotating between two sets of fixed vanes U. The difference between maximum and minimum capacitance is 64 pF.

Connected parallel to this capacitor is a second one with a constant capacitance, such that the total capacitance in the minimum position of the rotating capacitor is approximately the desired value $C_2 =$ 400 pF. This fixed capacitor can be slightly adjusted with the aid of a servomotor, during operation, to get exact resonance between the frequency of the oscillator and the travelling particles in the accelerating chamber. Resonance can be judged by the occurrence of a maximum in the current flowing to the target, or by the concentration of neutrons around the accelerating chamber.

Both the rotating capacitor and the fixed capacitor are mounted in an evacuated drum (10^{-5} mm Hg), thus making it possible for the clearances between the rotating and the fixed capacitor parts to be kept relatively small (6 mm) in spite of the high voltages prevailing, and thus to obtain the required large capacitances with a comparatively small plate surface. In this way it has been possible to keep the modulator small enough to be able to work with the high speed of 3000 r.p.m. and thus reach the high modulating frequency of 2000 c/s [1]).

[1]) In this connection it is important that a capacitance variation of only 64 pF is needed. If the frequency sweep has to be much greater than 4% (thus for a higher final energy of the particles, and for protons instead of deuterons; see I), then rotating capacitors of a much higher value are needed, and it is then not so easy to reach high modulating frequencies.

From the circuit diagram of *fig. 8* it is seen that one plate of the fixed capacitor (C'') and the fixed vanes of the variable capacitor (C') are connected to the inner conductor of the transmission line. The other capacitor plate and the rotating toothed disc have to be earthed. The earthing of the rotating disc could be achieved, in principle, by means of a sliding contact (brushes), but this was found to be unsatisfactory owing to the too high brush resistance for high-frequency currents and on account of arcing arising from the sparking in vacuo. The disc is therefore earthed by means of a large capacitor consisting of a number of fixed, earthed plates and, rotating between them (with a spacing of 2 mm), a number of plates mounted on and electrically connected to the spindle of the rotating capacitor disc C'. The earthing of C' obtained in this way is not, it is true, complete, since the capacitances of C' in series with those of the earthing capacitor form a voltage divider, so that the rotating disc and the rotating plates of the earthing capacitor continue to be at a small high-frequency voltage, of about 0.5 kV. This necessitated insulation of the rotating part of the modulator in the manner indicated in *fig. 9*. The disc and the plates are mounted on a bush turning in two ball bearings about a fixed, water-cooled, shaft insulated with blocks of a ceramic insulating material ("Kersima"). The set-up is such that the capacitance between the shaft and earth is small, so as to avoid any appreciable high-frequency currents flowing across the ball bearings.

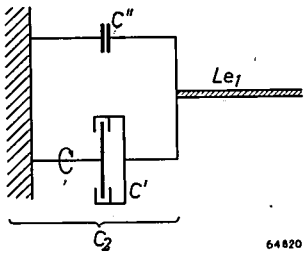


Fig. 8. Circuit of the modulator. The periodically varying capacitor C' is shunted by a constant capacitor C'' for making up the desired total capacitance. The rotating disc of C' has to be earthed.

These currents could be further reduced in a well-known way, by shunting the capacitance just mentioned by a coil to form a circuit tuned to 10.7 Mc/s, which thus has a maximum impedance for that frequency.

The rotating part of the modulator is driven by a motor also placed in vacuo. It would have been possible to drive the modulator from outside the

vacuum drum but then it would have meant having to pass through the wall of the drum a motor shaft continuously rotating at a speed of 3000 r.p.m. while still maintaining a vacuum of 10^{-5} mm Hg, which would not have been so easy to ensure.

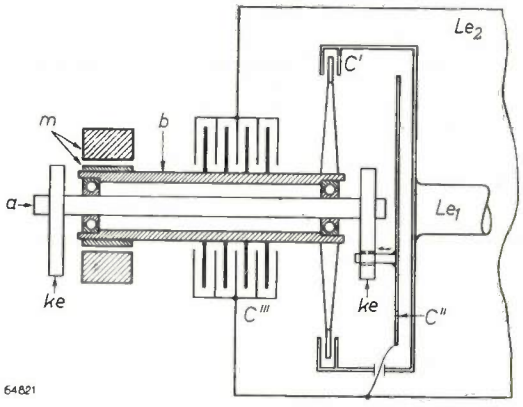


Fig. 9. Construction of the modulator, with the rotating capacitor C', the fixed (adjustable) capacitor C'' and the earthing capacitor C''' (schematically drawn). Le_1 and Le_2 inner and outer conductors of the transmission line; b bush carrying the rotating capacitor parts and rotating about the shaft a insulated with two sets of "Kersima" blocks ke; also mounted on b is the rotor (an iron bush) of the induction motor m.

Our solution is very simple, only the three-phase voltage supply cables for the motor having now to be led in through the wall of the drum. The rotor of the induction motor has the form of an iron bush slipped onto the bush mentioned above, which carries the rotating parts of the modulator. The three-phase stator producing the rotating field, is placed around the bush with a gap of a few mm in between. An essential condition is that in vacuo the motor must not give off too much gas or impurities, and this condition is very difficult to fulfil with field coils of insulated wire. The motor has therefore been built for a very strong current (each of the three phases 185 A, voltage 2.1 V; the starting current is 300 A) and a very small number of windings. Thus the windings could be made in the form of free, bare copper tubes through which cooling water can flow and which are mounted on two "Pertinax" plates (the only insulating material contained in the motor). This and other details of the construction are shown in *fig. 10*.

As the motor has only a small torque and the system has to start quickly in order to pass smoothly over some critical speeds, the bearings of the system have been very carefully designed. The friction is so small that it takes more than an hour for the system to run out to a standstill (it is to be remembered that there is no air resistance). The bearings can be lubricated from the outside with special vacuum oil, but this proves to be hardly ever necessary.

It has been mentioned that the dimensions of the parts of the oscillator, including the modulator, had to be determined experimentally with models in order to make sure of getting the right frequency. Measurements taken show that as a result the frequency fluctuates between 10.46 and 10.11 Mc/s. This is a variation of slightly less than 4%, but it proved to be more than sufficient. From the resonance condition $\omega = qB/m$ it follows that at the maximum frequency resonance occurs when $B = 1.36$ Wb/m^2. The actual flux density in the middle of the accelerating chamber is a little more than this, so that resonance is not reached until the frequency has dropped slightly. Consequently during a cycle of the modulating frequency acceleration takes place roughly between the instants M and N in *fig. 11*, i.e. during somewhat less than half 1/2000 sec, thus about 1/5000 sec.

The shape of the frequency variation curve, which is related to the chosen form of the teeth and blades of the rotating capacitor, is not critical, owing to the phase stability of the travelling particles.

Here something has to be added about the modulating frequency. It has been pointed out that the high value of 2000 c/s is favourable for obtaining a high average beam current on the target. The

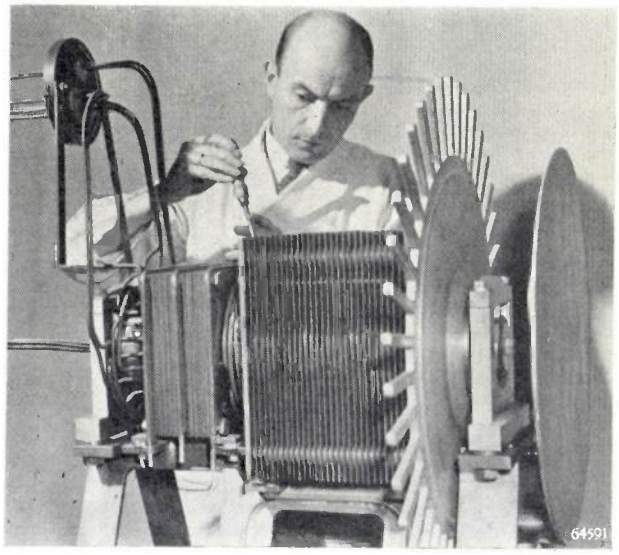


Fig. 10. The modulator (without the parts of the capacitors C' and C'' fixed to the transmission line; see fig. 9). From right to left: the slightly adjustable plate of C'' (guided by three pins), the toothed disc of C', the large earthing capacitor C''', and the stator coil of the driving motor with the peculiar "windings" for the three phases made up of copper water pipes standing free.

question arises whether a still higher modulating frequency would be desirable.

In our synchrocyclotron the deuterons have a peripheral rotational speed of roughly 10^7 "loops" per second (viz. about equal to the frequency of 10.7 Mc/s of the voltage on the dees). The acceleration of a particle to the final energy of 28 MeV has to take place in about 1/5000 sec, in which time the particle describes $10^7/5000 = 2000$ loops, so that it is accelerated by the dee voltage 4000 times and has to gain 7 keV in energy each time. At a peak value of 14 kV of the sinusoidally alternating dee voltage, the particles coming under the phase-stabilizing action and accelerated to the final energy are those which pass the gap with a phase angle $\varphi = \arccos 7/14 \approx 60°$ with respect to the peak of the dee voltage (see A). If a higher modulating frequency is chosen then a smaller angle φ is needed to allow the particle to reach the final energy of 28 MeV in the shorter time available. This, however, makes the orbit less stable, the phase tolerances become smaller and there are fewer ions in a group. Consequently the mean beam current will increase less than proportionately with the number of current surges per second and ultimately even decrease.

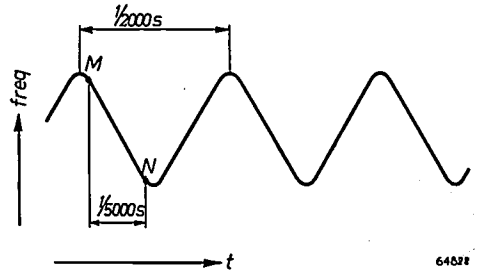


Fig. 11. Variation of the oscillator frequency with time. Modulation frequency 2000 c/s. Acceleration of the deuterons takes place during the interval *MN* of each modulation cycle.

Indeed it is seen that during the starting period of the modulator the mean current on the target increases first roughly in proportion to the modulation frequency and then more slowly. At 2000 c/s (and 14 kV dee voltage) the expected maximum is not yet reached. Constructional difficulties have for the time being prevented any further increase of the modulation frequency. It is particularly the mechanical stresses in the rotating toothed disc that then constitute a serious problem: at a peripheral speed of 110 m/sec the material has to withstand very great forces. That is also why, instead of copper, duraluminium was chosen for this disc.

We can now also understand the influence of the dee voltage on the choice of modulation frequency as already referred to in passing. If instead of 7 kV the gain in energy per "loop" were to be only 3 kV then the acceleration to the final energy would take $^7/_3$ times as long and the modulation frequency (for the same φ) could not be more than 850 c/s. The Harwell synchrocyclotron does in fact work with 3 kV gain per loop, so that for a final energy of 180 MeV the particles have to make 30,000 loops, and, as the mean oscillator frequency of that cyclotron is about 23 Mc/s, the modulation frequency cannot be made any higher than about 300 c/s.

Set-up of the transmission line with modulator and dee

Taking a final look over the whole assembly of oscillatory circuit with dee and modulator, we find there are still some other constructional problems typical for the synchrocyclotron.

The inner conductor of the transmission line must be insulated. In order to avoid excessive dielectric losses in the supporting insulators the latter have preferably to be placed where no high-frequency voltages occur. As we have seen in fig. 1, this is the case near the middle of the line. Thus, roughly speaking, we get the picture of a copper pipe about 15 cm in diameter and about 5 metres long supported only over a short distance near the middle, carrying at one end the weight of the dee and at the other end the vanes of the variable capacitor between which the toothed disc is rotating at a speed of 3000 r.p.m.

Since the copper pipe would not be able to bear this mechanical load it has been provided with an internal frame of brass.

Another complication is the fact that the pipe, at least at the ends, has to be in vacuo. The simplest solution would appear to be an arrangement whereby the whole transmission line is in vacuo, thus joining up the modulator drum and the accelerating chamber into one continuous vacuum. In that case the ports through which the coupling loops of the oscillator are passed would have to be vacuum-tight, whilst in the event of any changes having to be made to these couplings or in the rather complicated construction of the modulator the vacuum of the whole cyclotron would of necessity be disturbed and have to be restored. We therefore preferred to separate the vacuum of the accelerating chamber and the modulator and to give each its own pumping system (see I, fig. 1). The two oscillator coupling loops are now easily accessible and come to lie in air. On either side of the anode coupling loop are the supporting insulators carrying the inner conductor of the transmission line.

Since this conductor has to be led into the two vacuum chambers via vacuum-tight insulators it

would have been obvious to use the supporting insulators for this purpose. In fact, also the port insulators must not be too far away from the voltage node. Accordingly, the vacuum chamber of the modulator and also that of the accelerating chamber is in fact extended close to the voltage node by a cylinder enveloping the copper outer conductor of the line. Nevertheless it was found necessary to keep the functions of supporting and port insulators separate. These insulators already have to answer rather exceptional requirements: the dielectric losses at a frequency of 10 Mc/s have to be low, in connection with which also the dielectric constant has to be small, and they must stand up to a voltage of at least several kilovolts (they are still that far away from the actual voltage node). Furthermore, the port insulators have to make a vacuum-tight seal of rather large dimensions (the pipe to be passed through is 16 cm in diameter), and the supporting insulators have to be strong enough to bear the weight of the pipe with dee and modulator vanes. Only with the functions separated was there enough freedom in the shaping of the insulators for answering the requirements in both cases. As insulating material "Kersima" [2]) was again used, a ceramic material with very good electrical and mechanical properties, in particular exceptionally high compressive strength (porcelain would be quite unsuitable). *Fig. 12* shows the ultimate set-up.

To complete the picture it is to be mentioned that in the conductors of the transmission line, with dee and modulator, with high-frequency currents of about 250 A_{rms} in permanent working a power of over 3 kW is dissipated. Owing to the insulated mounting and the fact that the greater part of the system is in vacuo, only a small part of this heat can be carried off by conductance or by convection. Neither can radiation help much in cooling, since owing to the high frequency the surface cannot have a properly radiating layer but rather has to be bright. This accounts for the extensive use of water cooling, which to the uninitiated may at first sight be astonishing. Water pipes run to all parts of the transmission line, the dee and the modulator. Of course the cooling water inlet is in the middle of the line, at the point of the voltage node. Since, however, in connection with the frequency modulation this node swings somewhat to and fro, the pipe system is connected to the municipal water mains with insulating hoses of polystyrene rubber (see fig. 6 at the bottom on the right), which have only very low dielectric losses.

Finally, by way of demonstrating with an example the very close relationship of all constructional details of the synchroncyclotron, it may be pointed out that it is in fact due to the properties of the ceramic insulating material that two transmitting valves are needed in the oscillator: if the properties of the insulating material were such that the supporting and port insulators could be placed farther away from the middle then more space would be

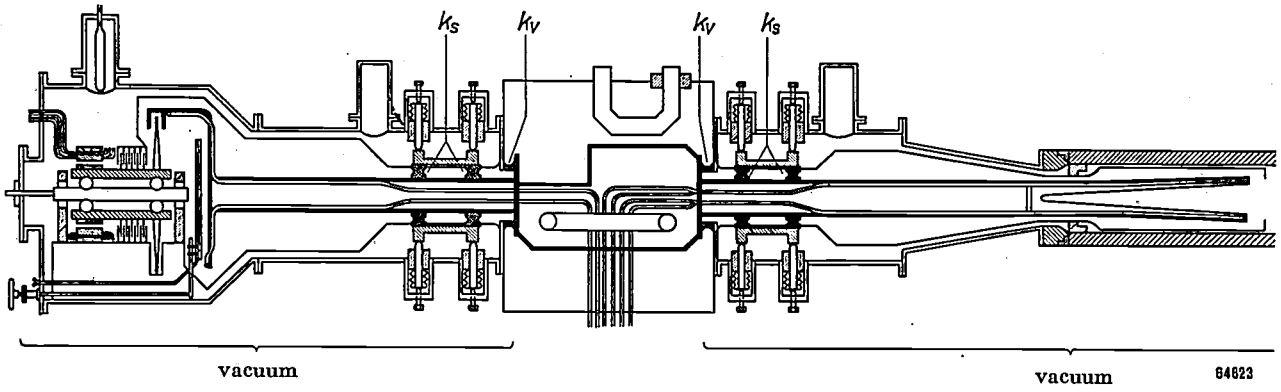


Fig. 12. Set-up of the oscillatory circuit. The approx. 5 m long copper inner conductor of the transmission line, with at the ends the dee and the vanes of the modulator (all drawn in heavy lines), is supported by the ceramic insulators k_s, which, in order to reduce their electric load, are placed not far from the voltage node in the middle of the line. To give the wide copper pipe the necessary strength it is fitted with a brass frame on the inside (not show in the drawing). For centering the system the supporting insulators can be shifted by means of screws (and vacuum-tight metal bellows). The dee and the modulator are each in separate vacuum chambers, so that the middle part of the line with the coupling loops is readily accessible. The ceramic insulators k_v provide for vacuum-tight ports through which the line can be passed. Also part of the cooling water pipes is shown in the drawing.

[2]) Cf. Philips Techn. Rev. **10**, 205, 1948.

available for the anode coupling loop, higher values of the inductances M_1 and M_2 could be reached and the oscillatory condition (2) could be fulfilled with a smaller mutual conductance, thus with only one valve.

Summary. I. The Philips synchrocyclotron at Amsterdam is producing in continuous operation deuterons with an energy of 28 million electronvolts, which are being used for many forms of nuclear transmutations. The fundamental considerations to be met in the designing of the installation were, apart from the choice of the particles to be accelerated and the final energy desired, that the beam current should be as high as possible and the working as reliable as possible. At this moment a current of 20 microamperes is already being reached with the energy mentioned. This energy of 28 MeV may just lie within the scope of the classical cyclotron, but the application of the principle of the synchrocyclotron makes the construction very much simpler and the installation more economical in working. In the general description the most important parts of the installation are mentioned and an idea is given of the extent of the accessory apparatus required.

II. In a synchrocyclotron the particles to be accelerated can without objection be made to describe a large number of loops, so that in this case a dee voltage of only about 15 kV_{peak} is needed. For resonance with the travelling particles the frequency of this voltage has to be 10.7 Mc/s (wavelength about 30 m). The oscillatory circuit in which this frequency is generated consists of a coaxial transmission line with at each end a high capacitance (about 400 pF) formed by one of the dees (the other dee is earthed) and the modulator respectively. Owing to these large capacitances, high-frequency currents of about 350 A (peak value) are flowing in the system, but by very careful construction the dissipative resistance has been limited to about 0.05 ohm, so that the dissipation is no more than about 3 kW. Of this about 560 watts is dissipated as "useful energy" by the beam of deuterons on the target. All parts of the oscillatory circuit are cooled with water. The transmission line is made to oscillate by means of a Meissner circuit, comprising two TA 12/20 transmitting valves connected in parallel. The mode of vibration is such that a voltage node occurs in the middle of the line. Parasitic oscillations are suppressed by by-passing the grid of the valves to earth and giving the anode and cathode coupling loops a suitable configuration. The frequency is periodically varied by 4% by means of the modulator at the free end of the transmission line; the modulator comprises a capacitor with a rotating toothed disc in vacuo. The driving motor is likewise in vacuo, this being made possible by a special construction (few windings, field current 185 A, voltage 2.1 V). The modulating frequency is very high, viz. 2000 c/s. The vacuum spaces of modulator and accelerating chamber are separated. The insulated, about 15 cm thick, inner conductor of the transmission line is led into these spaces via port insulators made of "Kersima", which material is likewise used for the supporting insulators bearing the inner conductor, the (non-earthed) dee and part of the modulator. The supporting and the port insulators are situated close to the anode coupling loop, where the high-frequency voltage is low.

THE ELECTRO-ANALOGUE, AN APPARATUS FOR STUDYING REGULATING SYSTEMS

I. COMPONENTS AND FUNCTIONS

by J. M. L. JANSSEN and L. ENSING.

621-52:621.3.012.8:53.072.13

Mathematical operations such as adding, subtracting, multiplying, dividing, differentiating and integrating can be carried out with the aid of electrical circuit elements. On this principle it is possible to construct electrical models of, say, the processes of heat exchange or of mechanical processes. Such models are attractive because of the ease with which, by employing an oscilloscope, a sufficiently accurate picture can be obtained of the working of the process imitated, as also of the effect produced when various parameters are altered. Electrical models have proved their value particularly in the case of automatic controllers used to keep a certain working factor constant. An electro-analogue is a collection of instruments required for building up an electrical model for such cases and studying its behaviour. This first article will deal with the components required for building an electro-analogue. Details of the electrical circuits will be discussed in another article to follow.

In industry nowadays automatic controllers are being employed on a large scale for stabilizing certain quantities — temperature, voltage, current, resistance, acidity, rate of flow of a liquid or gas, etc. Some considerations were given to this subject in a previous article in this journal [1]), where it was pointed out that in a particular case, for designing the best regulating system or for the best manner of adjusting a particular automatic controller, the purely theoretical process of the calculation cannot be followed: the mathematical difficulties in taking into account all the pertinent factors would be far too great even in fairly simple cases. Neither is it advisable, in many cases, to follow entirely empirical methods, since this may lead to lengthy interruptions in the industrial process, with the resultant decline in the quantity and quality of the production.

As already mentioned in the article quoted, a solution can often be found by working with an electrical model of the process to be controlled and of the controller itself, thereby choosing a time scale such that the phenomena can be observed on the screen of an oscilloscope. Measurements of the actual process can be limited to the recording of the step-function response, i.e. the characteristic indicating how the quantity to be regulated responds as a function of time to a sudden change of the regulating quantity at the input of the process.

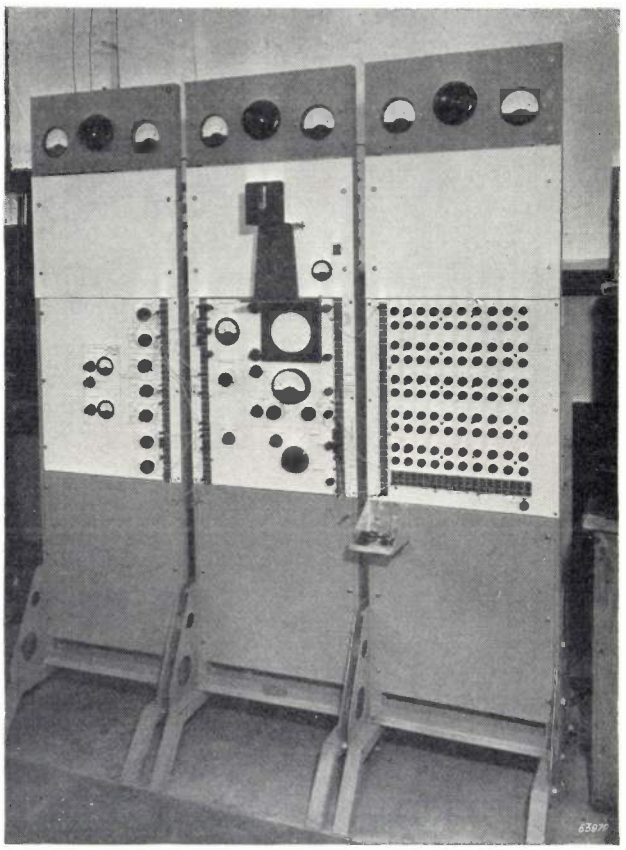


Fig. 1. Photograph showing the electro-analogue. The left-hand panel comprises mainly the model of the automatic controller, the middle one the oscilloscope and accessories, and the right-hand one the model of the process. At the top of each rack is a variable transformer for adjusting the A.C. voltage supply to the correct value, shown by the voltmeter on the right, and an ammeter (on the left) indicating the total current consumption of the panel.

[1]) H. J. Roosdorp, Philips Techn. Rev. **12**, 221-227, 1950/51 (No. 8).

The electrical model of the process has to be of such a construction as to show the same step-function response.

Given an apparatus which is designed for building up such an analogous model of the process and which comprises, further, a model of the automatic controller in which a proportional, an integrating and a differentiating term can be made to play a part as needed, and which moreover comprises an oscilloscope, a square-wave generator and further accessories, then the effect of altering various parameters in the controller, of introducing disturbances at various points, etc. can be seen at a glance. Such an apparatus is the "electro-analogue", a specific design of which is illustrated in *fig. 1*.

In addition to solving concrete regulating problems the electro-analogue has proved its worth in research work in the field of regulating. The great variety of models — both linear and non-linear — that can be built with it and the clarity of the results have contributed much towards a broader insight into regulating problems.

We shall first describe the components of which the electro-analogue built by us consists, while the electrical execution of the most interesting parts will be discussed in another article.

Process analogue

Analogous networks for some idealized processes

There has been no lack of attempts to put the step-function response of a more or less idealized process into mathematical form and then to translate that form into an electric network serving as a "process analogue", i.e. as an electrical model of the process.

Many processes are mainly characterized by an inertia caused, for instance, by a thermal resistance and a thermal capacity. Obviously in such cases the process analogue can to a first approximation be composed of a resistance R and a capacitance C in series (*fig. 2a*). The equation for the step-function response of this circuit (fig. 2*b*) is:

$$\frac{x}{x_0} = 1 - e^{-\frac{t}{RC}}.$$

where x_0 is the amplitude of the applied voltage step, x is the voltage across the capacitor and t is the time.

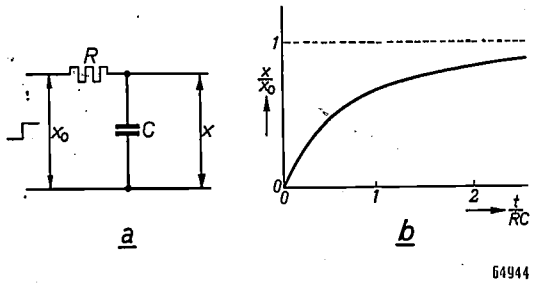


Fig. 2. *a*) Resistor R and capacitor C in series, as the simplest electrical model of a process with thermal resistance and thermal capacity. *b*) Step-function response curve of this network.

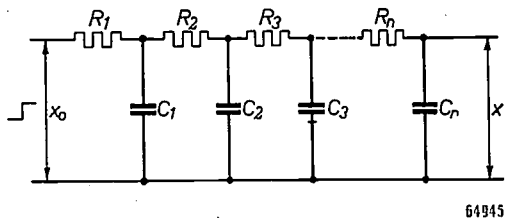


Fig. 3. Extension of the network in fig. 2*a* to a series of n R-C sections connected in cascade.

Usually, however, one has to do with a series of thermal (or other) resistances and capacities. A circuit according to *fig. 3* then gives a better approximation of the situation. In this network the sections R_2-C_2, R_3-C_3 etc. constitute a certain load on all the preceding sections, as is more or less the case in an actual process.

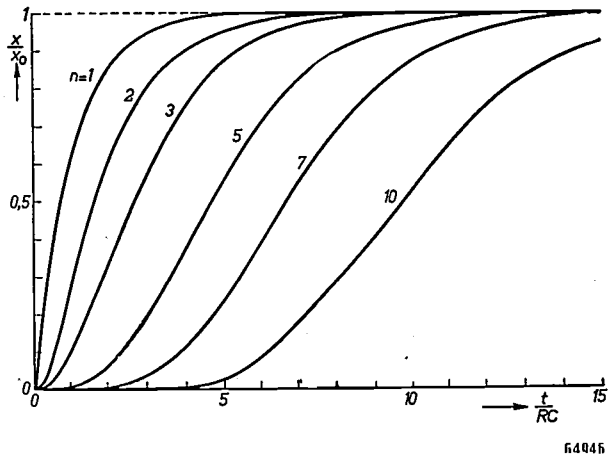


Fig. 4. Calculated step-function response curves for $n = 1$ to 10 sections of the network according to fig. 2 with equal resistors R and capacitors C in all sections, these sections being separated by interposed amplifying valves so that the one section does not constitute a load on the preceding sections.

Other hypothetical processes in combination with an automatic controller have been theoretically studied by mathematicians of the "Bataafsche Petroleum Maatschappij" [2]: one of these processes has its electrical analogue in a series of R-C sections where there is no such loading of the preceding sections by the following ones as just referred to.

[2]) P. Hazebroek and B. L. van der Waerden, The optimum adjustment of regulators, Trans. Amer. Soc. Mech. Engrs. 72, 317-322, 1950 (No. 3).

This case can be realized by separating the sections in the network of fig. 3 by means of amplifying valves (details will be given in a subsequent article). The calculated step-function response curves of such a network for different numbers of sections are given in *fig. 4*.

To verify the proper functioning of the electro-analogue this is provided with both kinds of *R-C* networks. The network without interposed valves consists of 20 "coupled", and that with valves of 40 "decoupled" identical sections, any number

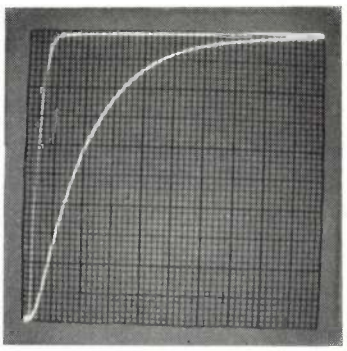


Fig. 5. Oscillograms of the step-function response curves of 10 sections of an *R-C* network; upper curve with decoupled sections, lower curve with the sections not decoupled. The upper curve agrees with the calculated curve for $n = 10$ in fig. 4, in which figure a widely different time scale has been used.

of which can be used as required. *Fig. 5* shows the oscillograms of the step-function responses of 10 sections of both networks.

These networks have rendered good service particularly for research purposes, but they do not form a universal process analogue with which any step-function response curve can be obtained. To arrive at such a universal process analogue, instead of adhering to an electrical analogy with the physical working of the actual process, an entirely different network has been devised by means of which any step-function response can be sufficiently approximated by a discontinuous line.

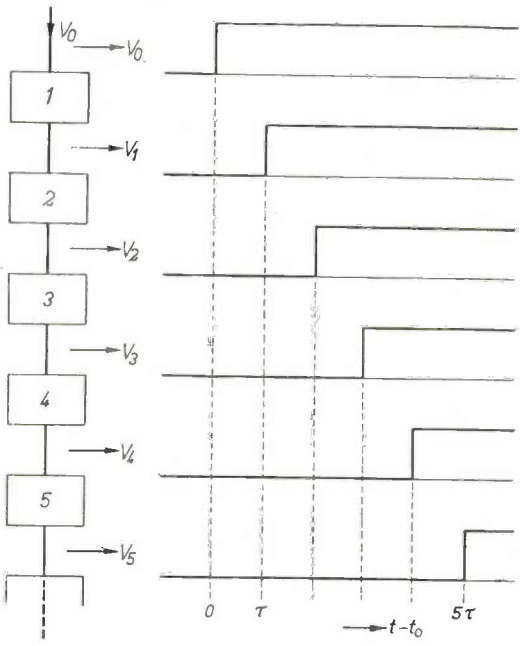
Universal process analogue

In the universal process analogue used by us an electric delay network [3]) is employed. This consists of a large number of sections, made up of capacitors and inductors, so arranged that voltages are transmitted, unattenuated, with a delay time τ which over a very wide frequency range is independent of the frequency. When a voltage V_0 having the shape of a step-function is applied to the input of the network then the output voltage V_1 of the first section likewise has the shape of a step-function but the step is delayed by a time τ (*fig. 6*) and is thus produced at $t_0 + \tau$; the voltage V_2 behind the second section, also with the shape of a step function, makes the step at $t_0 + 2\tau$, and so on.

If one now has a device with which any arbitrary proportions $a_1, a_2, a_3 \ldots$ of the voltages $V_0, V_1 V_2 \ldots$ can be added — which proportions it must be possible to adjust independently of each other between the limits $+a_{max}$ and $-a_{max}$ — then one can obtain a sum voltage V_s:

$$V_s = a_0 V_0 + a_1 V_1 + a_2 V_2 + \ldots,$$

of echelon shape as shown in *fig. 7*, and with this



64947

Fig. 6. Left: delay network with sections *1*, *2*, *3*, ... When a voltage step V_0 (see right-hand part of the diagram) is applied to the input then there arise at the outputs of the sections equally large pulses V_1, V_2, V_3 ... lagging a time τ, 2τ, 3τ, ... behind V_0.

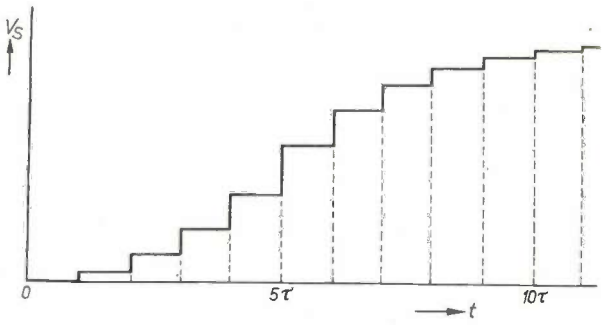


Fig. 7. When variable fractions a of the voltages V_0, V_1, V_2, V_3, ... of fig. 6 are added together, a sum voltage V_s of echelon shape is obtained, with which any curve can be approximated by varying the coefficients a.

[3]) British patent specification No. 517,516 in the name of A. D. Blumlein, H. E. Kallmann and W. S. Percival.

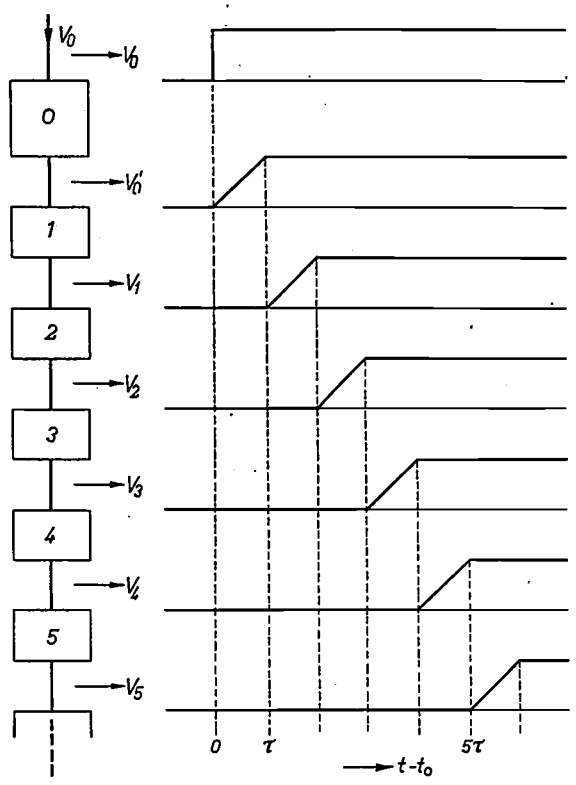


Fig. 8. As in fig. 6 but with the delay network preceded by a section *0* which converts the step voltage V_0 into a voltage V_0', which in the interval τ rises linearly with time.

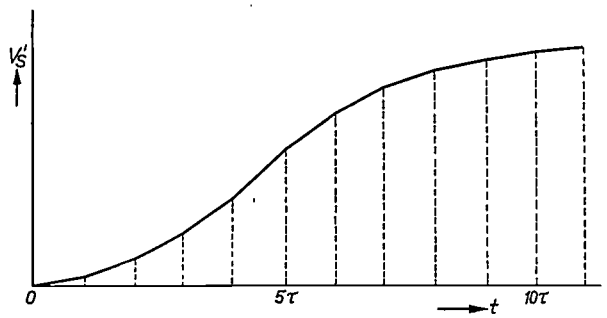


Fig. 9. Adding up variable fractions a of the voltages V_0', V_1, V_2, V_3, ... of fig. 8 yields a sum voltage V_s', the curve of which more closely approximates to a smooth line than the curve of fig. 7.

it is possible to approximate quite well the measured step-function response of any actual process.

The approximation is much better still when the voltage V_0, having the shape of a step function, is first converted into a voltage V_0' which in the interval τ increases, for example linearly, from 0 to the maximum value. This conversion can be brought about in a section (*0* in *fig. 8*) preceding the delay network. The shape of the voltages V_0', V_1, V_2 ... is then as shown on the right in fig. 8, and the sum voltage V_s' follows an almost smooth curve (*fig. 9*).

In our electro-analogue the method last outlined is applied. The delay network consists of 50 sections. There is no trace of any kinking in the curve of the sum voltage.

Oscilloscope and square-wave generator

For making the step-function response of the process analogue identical with that of the actual process it is necessary to be able to observe both these characteristics simultaneously. To that end the variation of the output voltage V_s' of the delay network is displayed on the screen of a cathode-ray oscilloscope, over which a transparent sheet of cellophane marked with a system of coordinates is placed, on which the measured step-function response of the actual process has been plotted. The curve of the process analogue is then matched as closely as possible to the plotted curve by turning the knobs with which the coefficients a_0, a_1 ... a_{50} are varied (see the right-hand rack in fig. 1).

For oscilloscopic observation it is advisable that the voltage V_0 does not make one single step but that this is repeated periodically, thus using a square-wave voltage (*fig. 10a*). This voltage is produced in a square-wave generator, in which the peaks of a sinusoidal voltage (50 c/s) taken from the mains are clipped off, leaving a practically pure square-wave voltage.

The output voltage V_s' of the process analogue now has a shape as represented in fig. 10*b*. The dotted part of this curve is made invisible by periodically blacking out the electron beam in the cathode-ray tube. The fully-drawn part is traced faithfully on the screen of the oscilloscope on account of the fact that in the respective half-cycles the time-base voltage is linear.

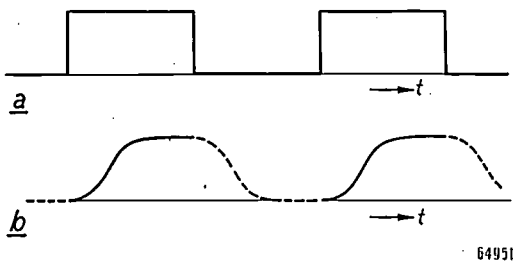


Fig. 10. *a*) Square-wave voltage at the input of the section *0* (fig. 8). *b*) Example of the variation of the sum voltage V_s' traced on the screen of an oscilloscope (the dotted part is blacked out).

Models of continuously acting automatic controllers

In the article quoted in footnote [1]) a distinction was made between discontinuously and continuously acting automatic controllers. We shall first consider only the latter type, leaving the other type

of controllers to be dealt with in the last section of this article.

In the equation for the relation existing between the amount x by which the quantity to be regulated deviates from the desired value and the amount q by which the regulating quantity is consequently

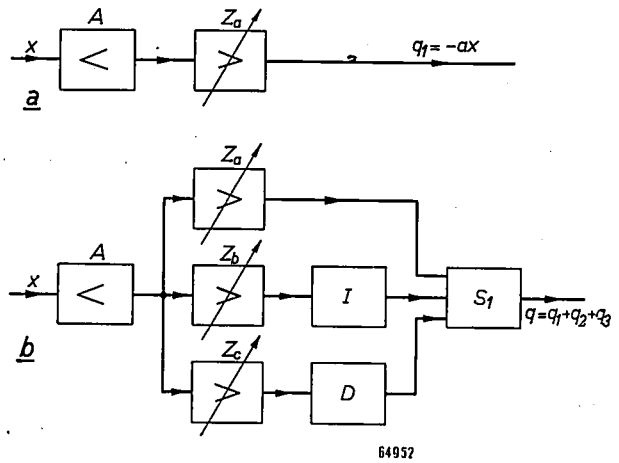


Fig. 11. *a*) Block diagram of a model of an automatic controller with proportional action: amplifier A and variable attenuator Z_a. *b*) Block diagram of the model of an automatic controller with proportional, integral and derivative action. A is a common amplifier, I the integrator, D the differentiator. The coefficients a, b and c of the three components are adjusted by means of the variable attenuators Z_a, Z_b and Z_c respectively. S_1 is a collecting stage in which the three components are added together.

changed by the automatic controller, various terms may occur, e.g. a proportional, an integrating and a differentiating term [4]):

$$q = -ax - b\int x\,\mathrm{d}t - c\frac{\mathrm{d}x}{\mathrm{d}t} = q_1 + q_2 + q_3. \quad (1)$$

It must be possible to realize these three terms in the electrical model of the automatic controller and the coefficients a, b and c have to be variable in the model.

The proportional term, $q_1 = -a\,x$, is easily realized by means of an electronic amplifier. The coefficient a can be made variable by adding a variable attenuator (*fig. 11a*).

The integrating term, $q_2 = -b\int x\,\mathrm{d}t$, and the differentitating term, $q_3 = -c\,\mathrm{d}x/\mathrm{d}t$, are obtained with the aid of an integrating and a differentiating network respectively. Separate attenuators provide for the variability of the coefficients b and c (fig. 11*b*). Finally the three voltages, q_1, q_2 and q_3, are summated in a "collecting stage".

[4]) In order to get positive numerical values of a, b and c we have given these factors a minus sign, in deviation from the article quoted in footnote [1]). That q and x must have different signs follows from the fact that the automatic controller has to bring about a change of q counteracting a preceding variation of x.

The electro-analogue is provided with a second integrator and a second differentiator by means of which terms of the type $b_2\int\int x\,\mathrm{d}t\,\mathrm{d}t$ and $c_2\,\mathrm{d}^2x/\mathrm{d}t^2$ can also be added, though as a rule there is no need for this.

Form and location of the disturbances

When a process analogue has been built up in the manner described then this network is combined with the model of the automatic controller according to fig. 11 *b* to form a closed circuit (*fig. 12*). Contrary to fig. 11*b*, a common attenuator Z_0 is added with which the coefficients a, b and c are varied simultaneously.

The oscilloscope can be connected to various points of the circuit by means of a lead. When it is found that the situation is stable — this can always be reached by adjusting Z_0 for sufficient attenuation — then at all input and output terminals of the component parts of the regulating circuit the voltage will be zero. Upon a disturbance being introduced somewhere, then at the output of the process analogue and at that of the model of the automatic controller voltages will arise which are proportional to the quantities x or q respectively, and on the screen of the oscilloscope it can be seen how these voltages vary with different values of the coefficients a, b and c.

The disturbances occurring in a regulating system in practice can be distinguished both ac-

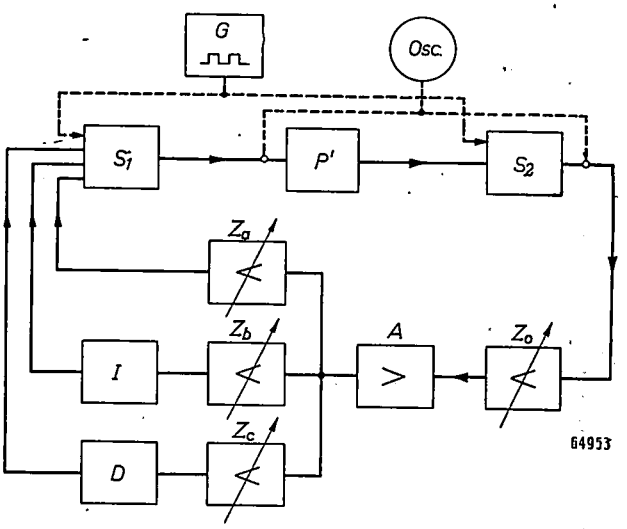


Fig. 12. Block diagram of the model of a regulating circuit. P' process analogue, whose step-function response has been matched to that of an actual process. Z_0, A, Z_a, Z_b, Z_c, I, D and S_1 form the model of the automatic controller (fig. 11). G square-wave generator supplying a disturbance, which can be applied, for instance, to the input of the process (at S_1) or at its output (at the collecting stage S_2 provided for that purpose). The oscilloscope *Osc* can be connected to various points of the circuit. Z_0 is a common attenuator.

cording to their form (as function of the time) and according to the point at which they occur in the circuit.

In respect to their form there are continuous and step disturbances. Since the former may be regarded as being the limit of an infinitely large number of infinitesimal steps occurring in succession (cf. fig. 6), it is only necessary to investigate the response of the system to step disturbances. These can therefore be obtained from the square-wave generator employed for adjusting the process analogue.

According to the point at which they occur, one may distinguish between:

1. disturbances in the regulating unit, thus at the input of the process,
2. disturbances somewhere in the course of the process, and
3. disturbances in the detecting element, thus at the output of the process.

There may also be counted among these "disturbances" the changes resulting from the automatic controller being adjusted to a different setting. It is with these distinctions in mind that the square-wave generator is also fitted with a lead for connecting it either to the input of the process (i.e. to the input of the collecting stage S_1) or to the output of the process, for which purpose a second collecting stage, S_2, is provided. If the disturbances introduced at these two points are sufficiently neutralized by the automatic controller then the same will be the case with the disturbances sub (2) occurring somewhere in the course of the process.

Oscillograms obtained with models of continuously acting automatic controllers

Some oscillograms will illustrate the influence of the various terms in the regulation equation (1). The lower curve in fig. 5 represents the step-function response of the process analogue used for the recording of these oscillograms (10 sections of the R-C network according to fig. 3).

First some examples are given of a regulating system with proportional action only, with $a = 4$ (12 dB) [5]). Fig. *13-I* shows that the process

[5]) In the model x and q are both voltages and therefore $a = -q/x$ is a non-dimensional quantity. But also in the actual process a is non-dimensional if x is expressed as a percentage of the desired value of the quantity to be regulated and q as a percentage of the corresponding value of the regulating quantity.

The practical advantage of expressing a in decibels is that the total value of a is found by adding up the readings of the various attenuators calibrated in dB (Z_0 and Z_a in fig. 12) and the gain (of A in fig. 12) expressed in dB, which is easier than multiplying.

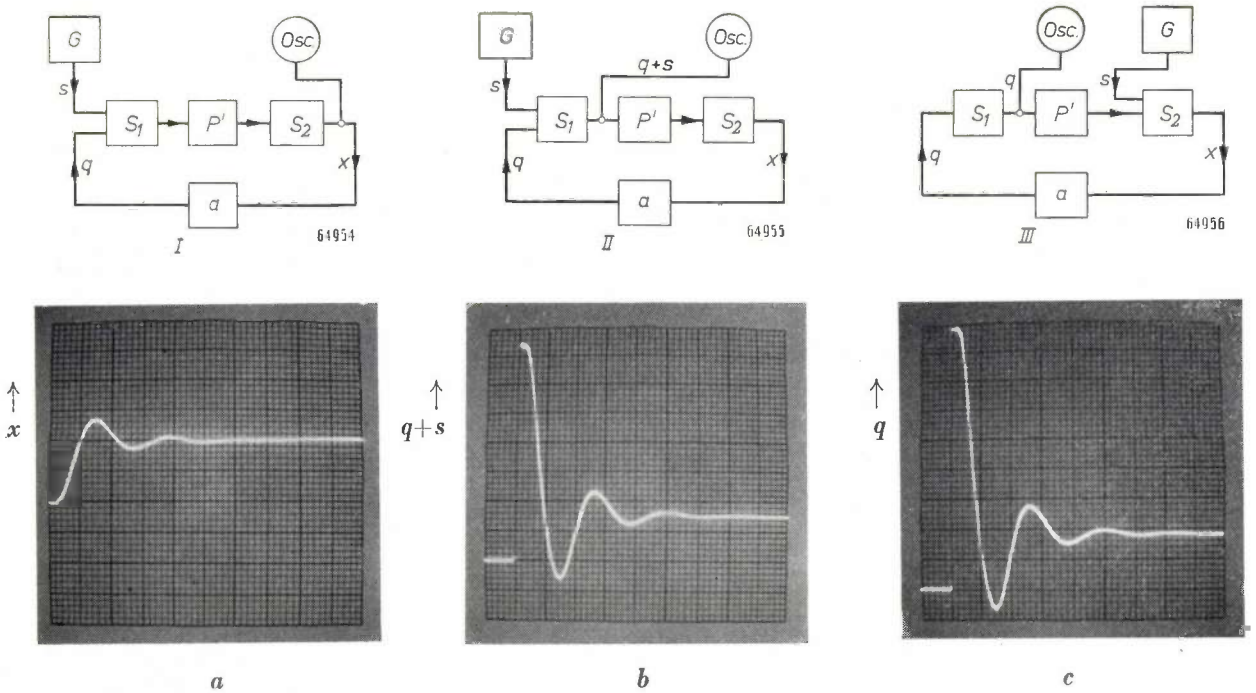


Fig. 13. Oscillograms obtained with the electro-analogue. The step-function response of the process analogue (10 not "decoupled" R-C sections) was as represented by the lower curve in fig. 5. The controller had only a proportional element ($a = 4$), as denoted by the rectangle a in the block diagrams *I*, *II* and *III*. In these diagrams it is also indicated at what points the generator G and the oscilloscope *Osc* were connected while the oscillograms underneath were being recorded.

analogue received, via the collecting stage S_1, the sum of the output voltage of the controller $q = -a\,x$ and the step-function voltage s from the square-wave generator, and that the output x from the process analogue was fed into the oscilloscope.

From the oscillograms it is seen that in course of time x and $q + s$ (thus q) become constant. The final values are related to the quantity a in the following way.

The gain of the process analogue (temporarily

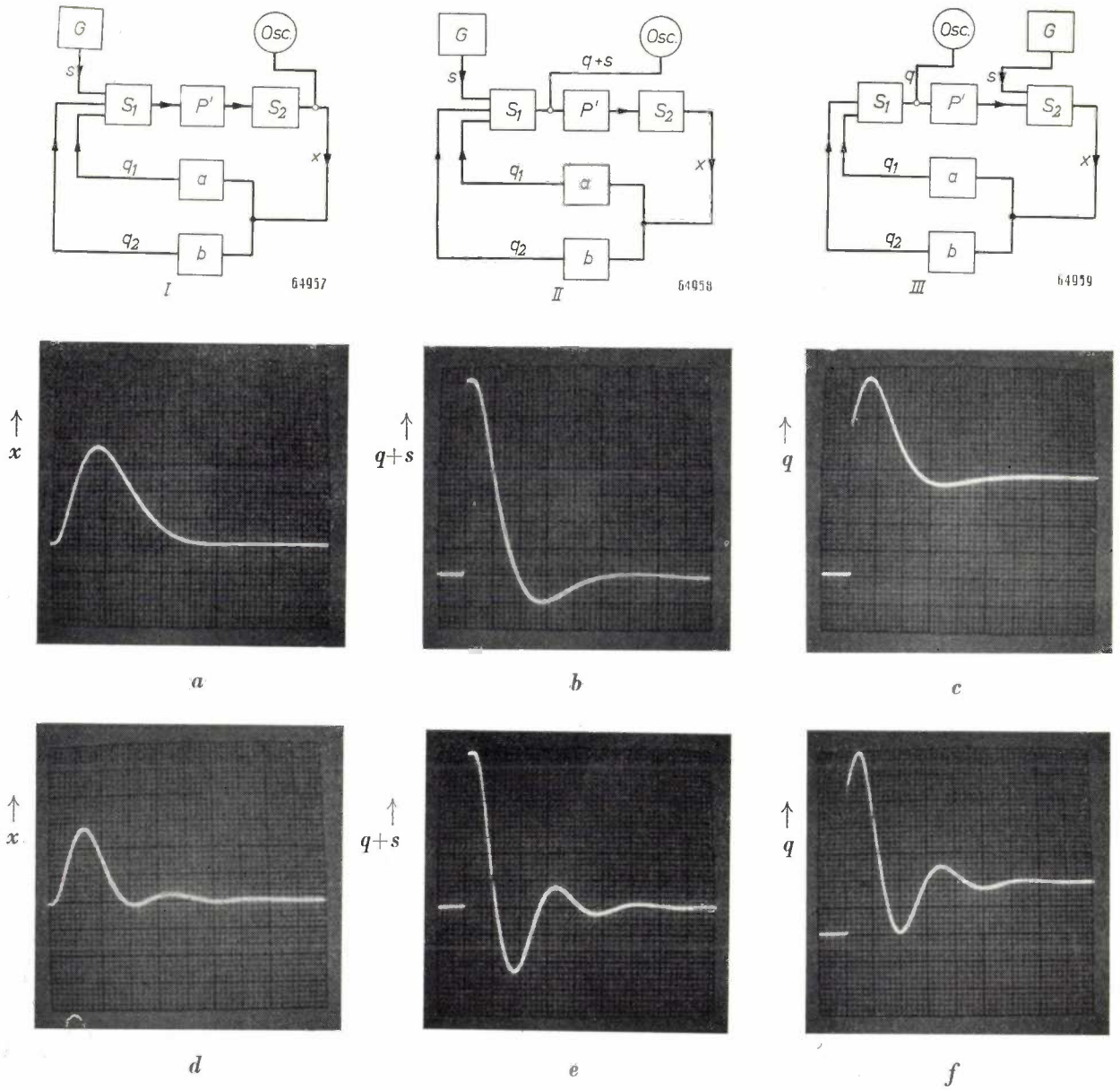


Fig. 14. As in fig. 13 but with an automatic controller having proportional and integral action. For the oscillograms (*a*), (*b*) and (*c*) the values of the coefficients were $a = 1.6$ and $b = 1.25$ millisec^{-1}, whilst for the oscillograms (*d*), (*e*) and (*f*) these were $a = 3.2$ and $b = 3.2$ millisec^{-1}.

Fig. 13*a* is the oscillogram obtained, on the same scale as fig. 5.

Fig. 13-*II* shows how $q + s$ is fed into the oscilloscope. In the corresponding oscillogram (fig. 13*b*) it is easy to distinguish the step-function voltage s from the square-wave generator, followed by the oscillation of q.

disconnected from the automatic controller) is so adjusted that a step disturbance with the amplitude s, introduced at the input, causes a change in the output voltage the final of value which, x_∞, is just equal to s. If, in the closed circuit, at the input of the process analogue we have in addition the output voltage q from the automatic controller

with the final value q_∞, then in the stationary state we find:

$$x_\infty = q_\infty + s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In the automatic controller the equation

$$q_\infty = -a\,x_\infty \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

holds. From these two equations it follows that

$$x_\infty = \frac{1}{1+a}\,s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and

$$q_\infty = \frac{-a}{1+a}\,s \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

From (4) it is seen that the larger the quantity a the smaller x_∞ becomes — i.e. the better a disturbance is neutralized — but that x_∞ never becomes exactly zero (offset). Equation (5) can be numerically verified in fig. 13*b*, for which $a = 4$.

According to fig. 13-*III*, in the collecting stage S_2 the step voltage s is added to the output voltage of the process analogue, the sum being termed x. Since the automatic controller contains only a proportional term the oscillogram of q (fig. 13*c*) is of the same shape as that in fig. 13*b*.

When the automatic controller contains an integrating element (*fig. 14*) then, as already explained in the article quoted in footnote [1]), x is indeed reduced exactly to zero after a disturbance has occurred. With the exception of the integrating element the networks *I*, *II* and *III* of fig. 14 correspond to those of fig. 13. The respective oscillograms in figs 14*a*, 14*b* and 14*c* relate to the case where $a = 1.6$ (4 dB) and $b = 1.25$ millisec^{-1}, while the oscillograms in figs 14*d*, 14*e* and 14*f* relate to the case where $a = 3.2$ (10 dB) and $b = 3.2$ millisec^{-1}.

The oscillograms in figs 14*c* and 14*f* can be interpreted as follows: owing to the inertia of the process analogue the prescence of this part is not yet noticeable immediately after the step. The oscilloscope thus shows a step —*as* (in the oscillogram a step upward). This is first followed by a linear rise, corresponding to the presence of the integrating element in the automatic controller, yielding a voltage $-b\int s\cdot dt = -b\,s\,t$ which increases linearly with time. Then, however, comes evidence of the presence

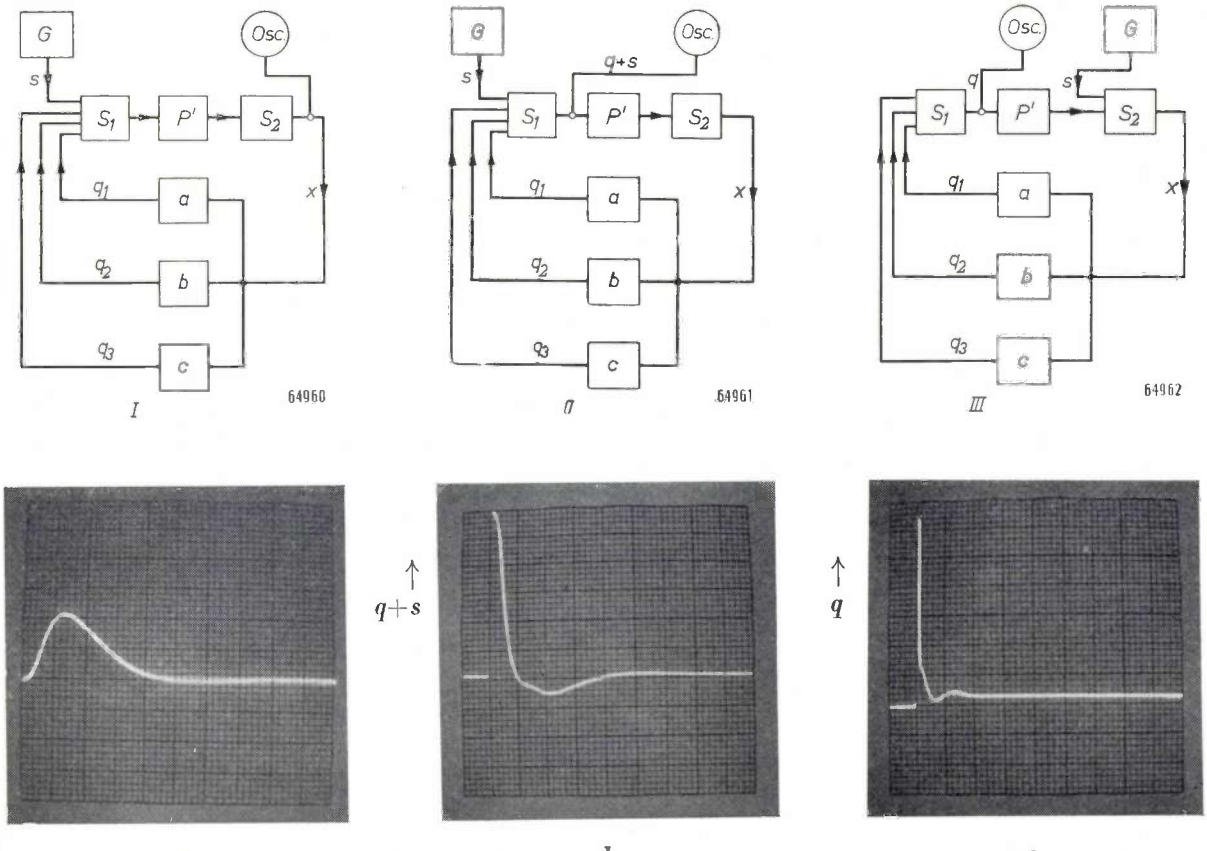


Fig. 15. As in fig. 14 but with also a derivative action in the regulator ($a = 3.2$, $b = 3.2$ millisec^{-1}, $c = 1$ millisec).

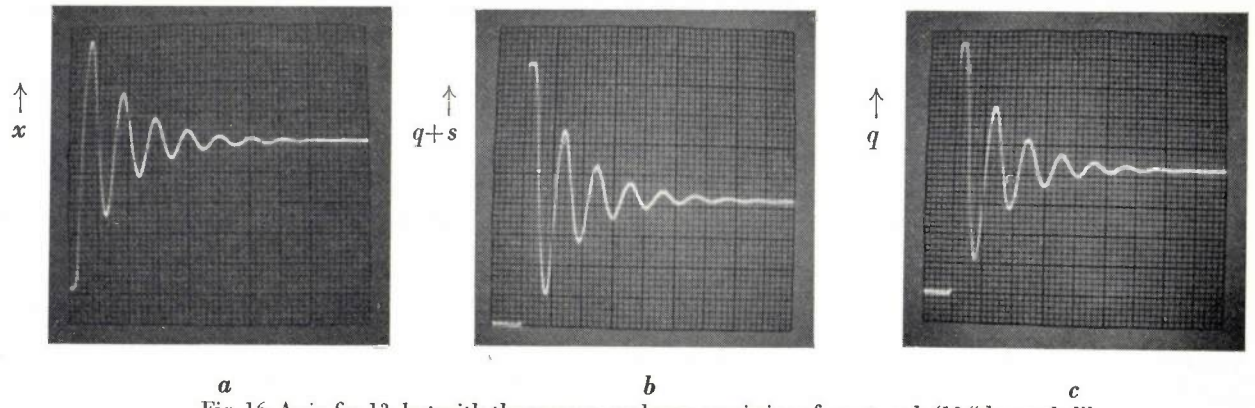


Fig. 16. As in fig. 13, but with the process analogue consisting of a network (10 "decoupled" *R-C* sections, with a step-function response as represented by the upper curve of fig. 5. Here the coefficient a of the proportional term was 1.

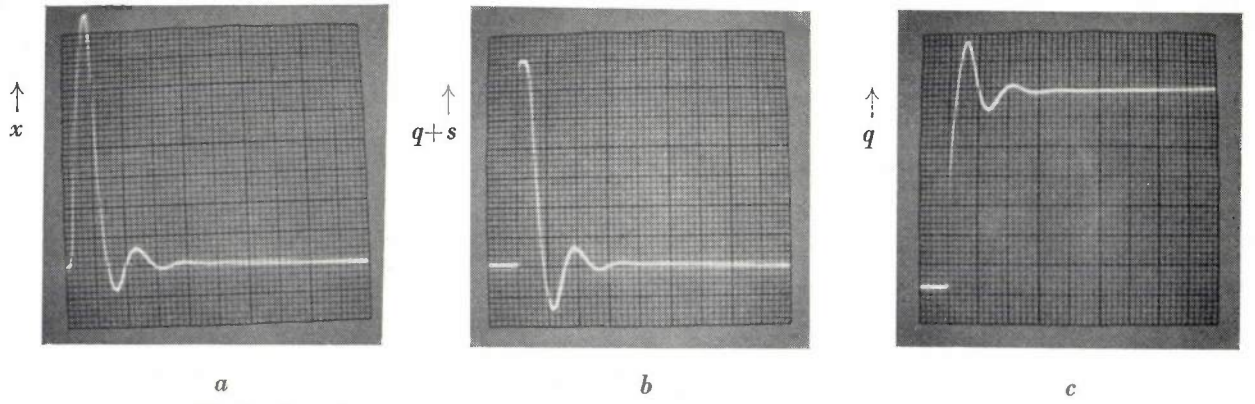


Fig. 17. As in fig. 14 but with a process analogue formed by a network with a step-function response as represented by the upper curve of fig. 5. $a = 0.5$,$b = 2$ millisec^{-1}.

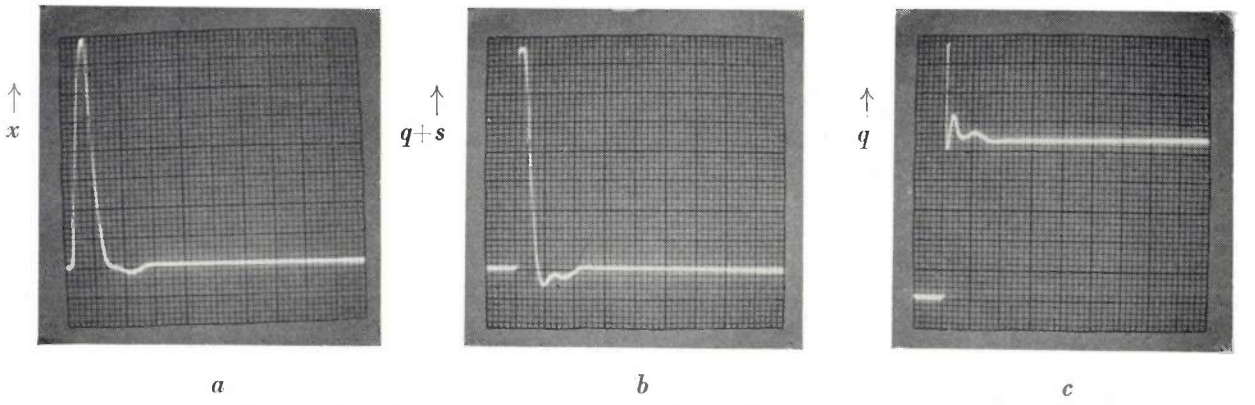


Fig. 18. As in fig. 15 but with a process analogue formed by a network with a step-function response as represented by the upper curve of fig. 5 and with coefficients $a = 0.5$, $b = 2.5$ millisec^{-1}, $c = 0.08$ millisec.

of the process analogue, which now also contributes a certain voltage to the input of the automatic controller. In the final state again equation (2) holds, but since now $x_\infty = 0$ we find $q_\infty = -s$, which means to say that in order to neutralize an interference at the output of the process a permanent change in the position of the regulating unit is necessary. Further the initial step q_0 made by q is equal to $-a\,s$. The ratio of this step to the final value of q, thus q_0/q_∞, is therefore equal to a, as may be seen from figs 14*c* anf 14*f*.

Finally in *fig. 15* similar oscillograms are shown for the case where the automatic controller contains also a differentiating element ($a = 3.2$, $b = 3.2$ millisec^{-1}, $c = 1$ millisec). Apparently the differentiation has a stabilizing effect upon the regulation.

Similar curves, but recorded with a process analogue with a step-function response as shown by the upper curve in fig. 5, are represented in *figs 16, 17* and *18* on the same scale as figs 13, 14 and 15. A comparison of these two groups of oscillograms shows that a process which can be represented by *R-C* sections loading the preceding sections can as a rule be better regulated than a (hypothetical) process corresponding to "decoupled" *R-C* sections; in the former case the regulated quantity gives smaller amplitudes (cf., for example, figs 13*a* and 16*a*), and the position in fig. 16, with a equal to only 1, is decidedly less stable than that in fig. 13 with $a = 4$.

Criteria for the best regulating system

Stability

An idea of the stability can be obtained at once from the oscillograms (e.g. figs 13 and 16): the less the damping of any oscillations present, the less is the stability. The curves do not show, however, whether, in a stable case, any small change in the working conditions or variation of the coefficients a, b and c may already result in instability. (Mechanical wear of parts of the regulating unit, for instance, may well cause a variation of the coefficients.)

A better insight into the degree of stability can be obtained by measuring the stability area. A process with only proportional action ($q_1 = -ax$) may become unstable when a exceeds a certain value. If the automatic controller has only an integrating component ($q_2 = -b\int x\,dt$) then stability may likewise be lost when b becomes too high. An automatic controller having both these functions has, in the case of a certain process, for each value of a a certain limit value of b above which instability occurs; the curve representing this limit value as a function of a forms, with the a-axis, the boundary of the stability area. If the automatic controller has also a differentiating action ($q_3 = -c\,dx/dt$) then for every value of the parameter c there is a different relation between the limit values of b and a. For the process analogue whose step-function response is represented by the lower curve in fig. 5, a number of curves determined with the electro-analogue with $b = f(a)$ and $c =$ constant are given in *fig. 19*.

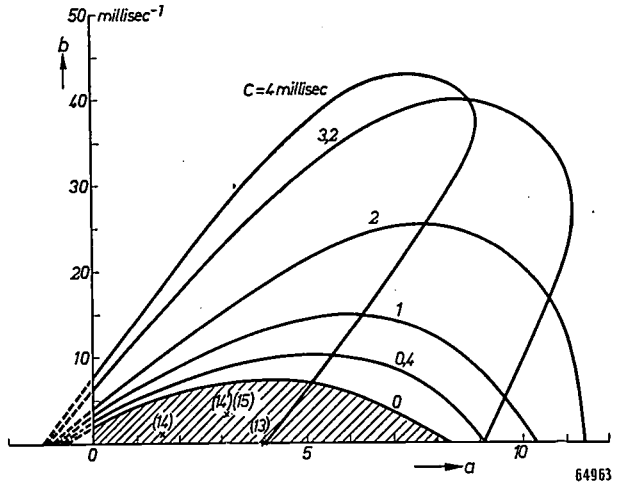


Fig. 19. Values of the coefficient b of the integrating term at which instability just arises, as function of the coefficient a of the proportional term, for constant values of the coefficient c of the differentiating term. This diagram applies for a process having a step function response as represented by the lower curve of fig. 5. The stability zone for $c = 0$ is hatched. The points (*13*), (*14*), (*14'*) and (*15*) correspond to the conditions under which the oscillograms of figures 13, 14*a*, *b*, *c*, 14*d*, *e*, *f* and 15 respectively were recorded.

If practical data are available concerning the variations likely to occur in the coefficients a, b and c, then with the aid of a number of graphs like those in fig. 19 it is possible to investigate whether there is any risk of the system becoming unstable.

Other criteria

Some valuable information is to be gathered from the oscillograms direct, as for instance the magnitude of the first maximum of x, the time that has to elapse after a disturbance before x remains within certain limits, the length of a cycle (if x is oscillatory, etc.

Other data are provided by the performance meter built into the electro-analogue, with which one can determine, as desired, the mean absolute value of x:

$$|\bar{x}| = \frac{1}{T}\int_0^T |x| \cdot dt,$$

or the r.m.s value of x:

$$x_{\mathrm{rms}} = \sqrt{\frac{1}{T}\int_0^T x^2\,dt}.$$

where T is the cycle of the repetition frequency, for which, as already stated, the mains frequency is chosen.

It may be required of an automatic controller that one of these quantities shall be as small as possible. It depends upon the case under investigation whether they are useful as criteria and, if so, which of them forms the best criterion. That is why both possibilities have been provided for in our electro-analogue.

Recording frequency response curves

In the modern theory of regulated systems use is often made of frequency response curves. The behaviour of the system as a whole and also that of its parts is studied with sinusoidal disturbances. Conclusions can be drawn therefrom as to the stability and the quality of the regulation. Literature on this subject has been published in abundance in recent years.

For sinusoidal disturbances to be introduced in the model of the regulating system the electro-analogue has been provided with an *R-C* oscillator the frequency of which is variable. This oscillator can be connected, for instance, to the collecting stage S_1 of the closed regulating circuit (*fig. 20*). In this way sinusoidal voltages arise at all points of the circuit. By measuring the input and output

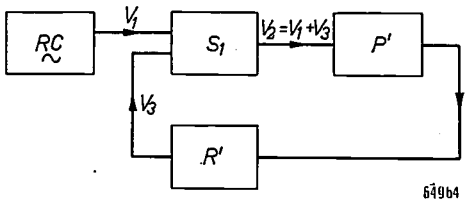


Fig. 20. Arrangement for taking Nyquist diagrams of the closed circuit of the process analogue P' and the model R' of the automatic controller. The (variable frequency) generator RC supplies a sinusoidal voltage V_1, which in the collecting stage S_1 is added to the output voltage V_3 from R'.

voltages of any element in the circuit one can determine the frequency response of that element. The same can be done for two or more elements connected in cascade, and even for the cascade connection of all the parts making up the circuit. In the latter case it is possible to derive from the measurements the Nyquist diagram, from which, as is known, conclusions can be drawn regarding the stability of the system [6]). Usually the measurements for the Nyquist diagram are taken after the circuit has been opened at some point. If, however, there is an integrator in the automatic controller (or the model) this is not possible, because when the circuit is cut open it becomes unstable: in fact the integral of an accidental step disturbance would continue to increase unrestricted, because in the opened circuit the disturbance is not neutralized. In that case the Nyquist criterion in its original form does not hold. In literature other forms of a wider scope are indicated [7]), but it would lead us too far afield to go into these here.

Models of discontinuously acting automatic controllers

Automatic controller with two-step action

The simplest form of a discontinuously acting automatic controller is that in which the regulating unit can have only two positions (as is the case, for instance, with a thermostat), so that the regulating quantity can assume only two values, Q_1 or Q_2. *Fig. 21a* represents the case — already discussed in the article quoted in footnote [1]) — where $Q = Q_1$ so long as the quantity X to be regulated is greater than a critical value X_{cr}, and $Q = Q_2$ so long as $X < X_{cr}$; usually it is so arranged that X_{cr} is just the desired value X_0. Between Q_1 and Q_2 is a value Q_0 at which — if Q could assume that value — under normal working conditions X would have just the desired value X_0. Instead of that, the regulating unit is alternately in the positions Q_1 and Q_2. If it is desired to change over to a different value of X_0 then Q_0 has to be given a different value too. If, however, Q_0 becomes greater than Q_2 or less than Q_1 then the automatic controller no longer functions.

In practice the transition from Q_1 to Q_2 will not usually take place at exactly the same value of X as the change-over from Q_2 to Q_1: owing to backlash in the mechanical parts of the regulating unit for instance, or owing to the difference between the current at which a relay is closed and that at which it is opened, the transitions take place at different values of X (X_1 and X_2) either side of X_0. Instead of fig. 21*a* we then get fig. 21*b*. Denoting the deviations of Q and X from the equilibrium values again by q and x, then $q = Q - Q_0$ and $x = X - X_0$; the regulating cycle $q = f(x)$ for an automatic controller with two-step action having backlash then has the shape shown in fig. 21*c*.

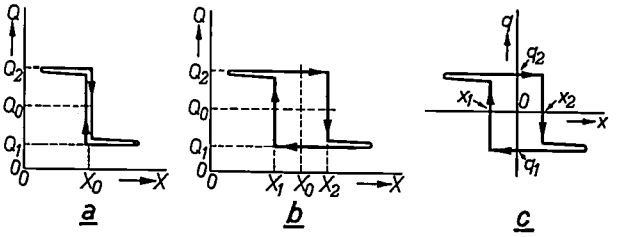


Fig. 21. Automatic controllers with two-step action, (*a*) without, (*b*) and (*c*) with backlash. In case (*a*) the regulating unit takes up the position corresponding to $Q = Q_1 < Q_0$ when X is greater than the desired value X_0, and $Q = Q_2 > Q_0$ when $X < X_0$ (Q_0 is the — impossible — position in which under normal working conditions X would assume the value X_0). *b*): The same but with a backlash X_1-X_2. In (*c*) the quantities $q = Q - Q_0$ and $x = X - X_0$ have been plotted against each other for the case (*b*).

The presence of backlash naturally reduces the response sensitivity and the accuracy of the regulation, but there is also a good side to it. In the case of an automatic controller without any backlash the regulating unit takes up the other position as soon as there is the least difference between X and X_{cr}, whereas in the case of a controller with backlash this does not take place until that difference reaches a certain value. In the latter case, therefore, the regulating unit is less subject to mechanical wear, and that is why sometimes some backlash is purposely given to the regulating unit.

It appears to be possible to build up electric circuits which show a relation between the output and the input voltages similar to that existing between q and x according to the hysteresis loop of fig. 21*c*. Such a circuit can serve as an electrical model for an automatic controller with two-step action. If, for instance, a sinusoidal voltage (x in

[6]) H. Nyquist, Regeneration theory, Bell Syst. techn. J. **11**, 126-147, 1932. See also B. D. H. Tellegen, Philips Techn. Rev. **2**, 292, 1937.

[7]) H. Bode, Network analysis and feedback amplifier design (Chapter VIII), New York 1945, and F. Strecker, Die elektrische Selbsterregung, Stuttgart 1947.

fig. 22) is applied to this circuit then a square-wave output voltage (q) is obtained.

In the electro-analogue containing such a circuit the latter can be combined with the process analogue to form a closed regulating circuit (*fig. 23*), in which also an amplifier, variable attenuators and a collecting stage are included. The limit $x_2 = -x_1$ is governed in the circuit by a direct voltage; thus by changing this voltage it is possible to vary the width of the loop. Via the collecting stage a direct voltage s can be added, corresponding to a variation of Q_0 or to bringing about a variation of Q_0.

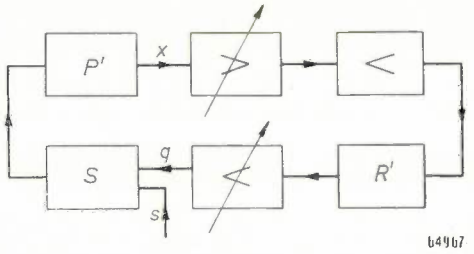


Fig. 23. Regulating circuit for studying discontinuously acting controlers (model R'). P' process analogue. Via the collecting stage S a direct voltage s can be applied to form the analogue of a disturbance or of a transition to some other desired value in the actual controller.

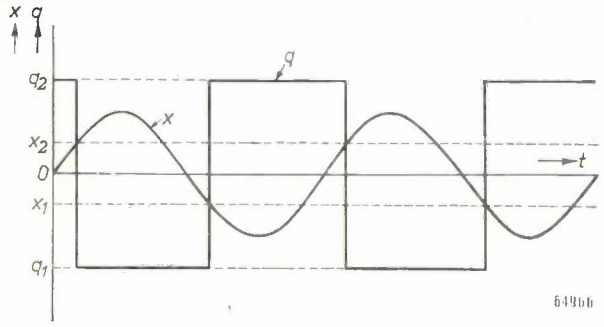


Fig. 22. When a sinusoidal voltage x (with amplitude greater than the limits $x_2 = -x_1$ of the backlash) is applied to an electric network forming the model of an automatic controller with two-step action there arises at the output of the network a square-wave voltage q.

In *figs 24a, b* and *c* oscillograms of q and x are represented respectively for $s = 0$, $s = +5$ V and $s = -5$ V, where changes takes place in the ratio of the intervals in which q has the value q_1 or q_2 respectively.

Automatic controller with three-step action

By giving the regulating unit a third position, Q_0, in between the positions Q_1 and Q_2 (*fig. 25*) one has the advantage that the system is at rest so long as the value of X is between X_1 and X_2.

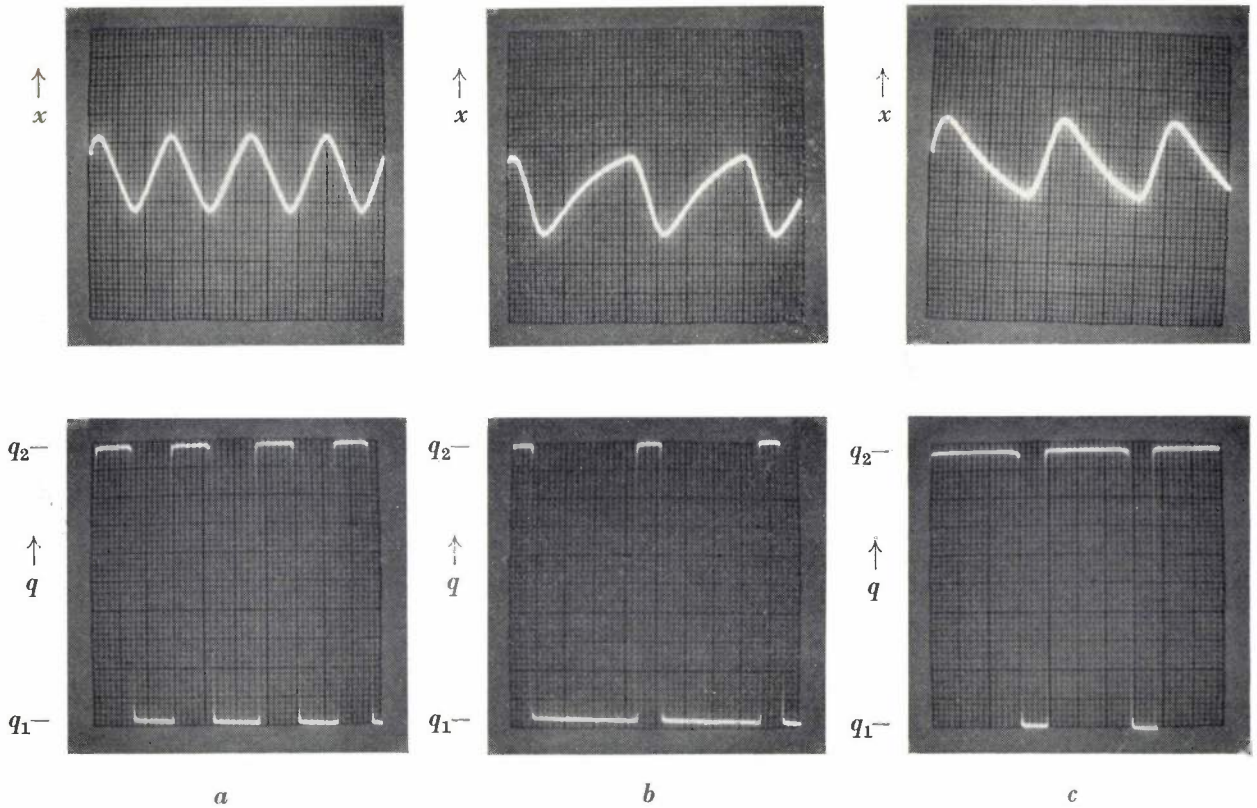


Fig. 24. Oscillograms obtained with a circuit according to fig. 23 with an automatic controller with two-step action. In the case (*a*) $s = 0$ and thus Q_0 lies halfway between Q_1 and Q_2 (fig. 21*b*). The intervals during which $q = q_1$ and q_2 respectively are of equal length. In the case (*b*) s is positive and in case (*c*) equally negative; the intervals during which $q = q_1$ and q_2 respectively are here unequal.

so that it suffers less from wear than a system lacking a state of rest. If considerable deviations from the normal working conditions arise only at long intervals then the controller and the regulating unit can remain at rest for a long time.

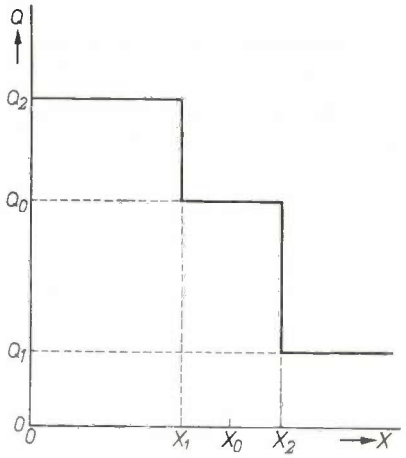


Fig. 25. Q as a function of X for an automatic controller with three step-action: in addition to the positions Q_1 and Q_2 also the intermediate position Q_0 is now possible. The regulating unit will be in the position Q_0 so long as X_0 remains between X_1 and X_2.

If, for some reason or other, X exceeds the value X_2 then the controlling unit causes the regulating unit to move to and fro between the position of rest Q_0 and the lower position Q_1, thereby correcting the too high value of X. The functioning can then be regarded as being similar to that of an automatic controller with two-step action (positions Q_0 and Q_1). If, on the other hand, X drops below X_1 then the controlling unit will fluctuate between the position Q_0 and the higher position Q_2, so that apparently we again have to do with an automatic controller with two-step action, but this time with the positions Q_0 and Q_1. Since the differences $Q_2 - Q_0$ and $Q_0 - Q_1$ are smaller than the difference $Q_2 - Q_1$ in the case of a controller with two-step action, also the periodic fluctuation of X caused by the changes of the controlling unit is smaller. This can also be regarded as an important advantage of an automatic controller with three-step action over one with two-step action. The fluctuation referred to can be reduced still further by increasing the number of positions of the controlling unit. The higher the number of positions the more closely the curve of echelon shape representing the relation between Q and X approaches a straight line, and the more the working of the automatic controller approaches that of a continuous controller with proportional action.

The functioning of an automatic controller with

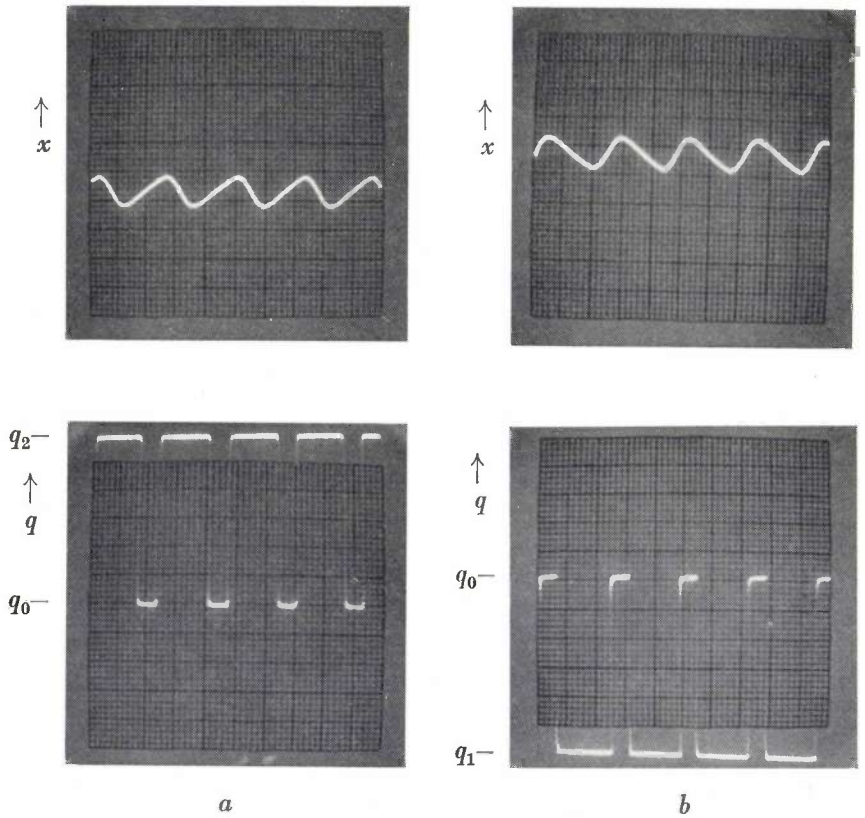


Fig. 26. Oscillograms obtained with a circuit according to fig. 23 with an automatic controller with three-step action, (*a*) with positive, (*b*) with equally negative disturbance voltage s.

three-step action is illustrated by the oscillograms in *fig. 26*. In this case, by applying a positive or a negative voltage s to the input of the process analogue, the rest level Q_0 was shifted in such a way that the controller came into action and moved to and fro between the positions Q_0 and Q_2, respectively between the positions Q_0 and Q_1.

Multi-speed floating control

Instead of the regulating unit being so made that it can only take up certain positions, it can be arranged so as to run with a certain, constant, speed dq/dt in the one direction or in the other or remain at rest according to the value of x. For instance:

$$dq/dt = -p \text{ when } x > x_2,$$

$$dq/dt = 0 \text{ when } x_2 > x > x_1,$$

and $$dq/dt = +p \text{ when } x < z_1.$$

This is made technically possible, for instance, by having a controlling unit with three-step action driving a servomotor which operates the regulating unit and turns at a constant speed in one direction or the other or remains at rest according to the position of the controlling unit. The motor then adjusts the regulating unit by an amount q varying linearly with the time: $q = \int p\,dt = p\,t + \text{const.}$

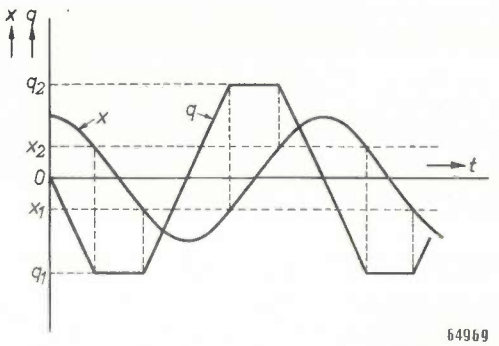


Fig. 27. When an alternating voltage x is applied to the input of an electric network forming the analogue of an automatic controller with three-step action the output voltage q assumes the shape of the drawn trapezium.

Compared with the automatic controller with three-step action, the multi-speed floating control has the advantage that in the event of permanent change from the normal (average) working conditions the motor provides for a permanent readjustment of the regulating unit.

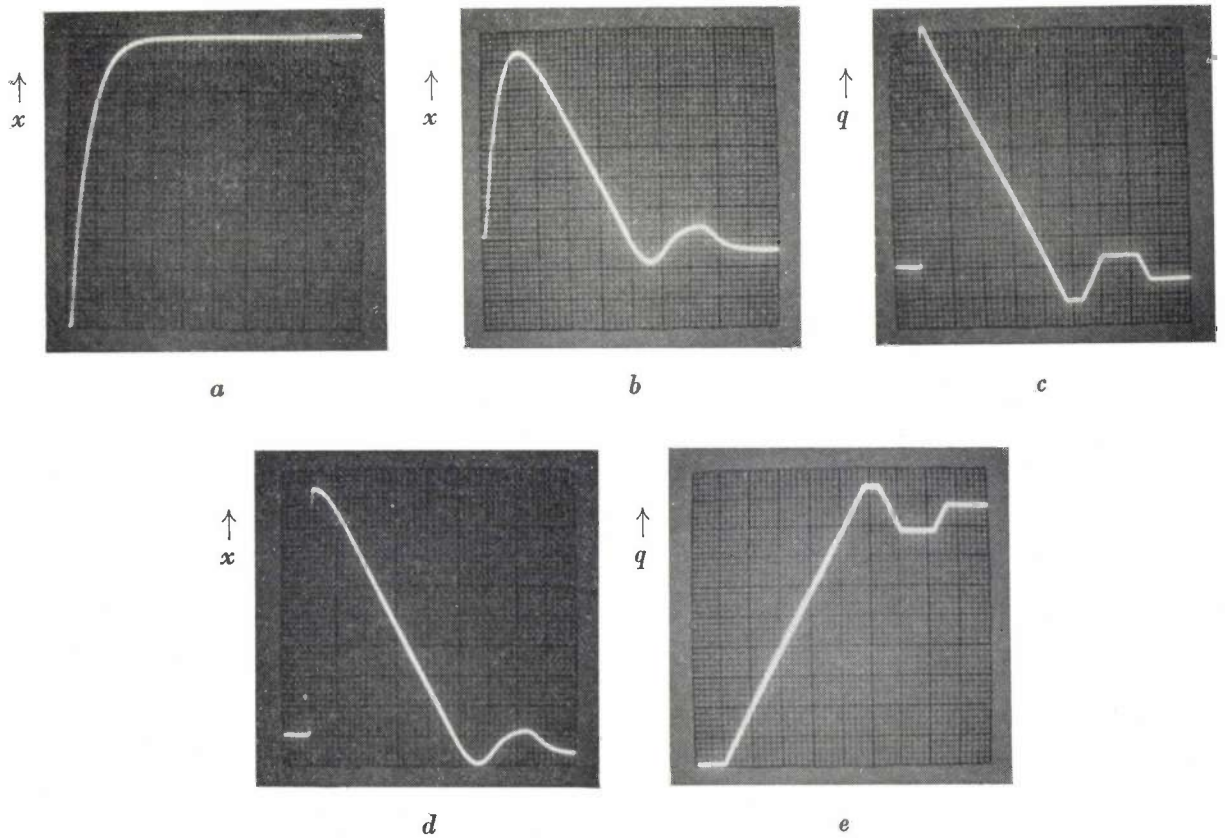


Fig. 28. *a*) Step-function response of a process analogue which together with an automatic controller with three-step action formed a regulating circuit from which the oscillograms (*b*) - (*e*) have been recorded. (*b*) and (*c*) represent the variation of x and q respectively in the case of a step disturbance at the input of the process, (*d*) and (*e*) the same quantities when a disturbance is applied at the output.

In an electrical model the function of the motor can be performed by an integrating network. When the output voltage ($-p$, $+p$ or 0) from a model of an automatic controller with three-step action is applied to an integrating network then at the output of that network there arises a voltage q which either increases or decreases linearly with the time or remains constant, thus in analogy with the readjustment of the regulating unit.

If, for instance, a sinusoidal voltage (with amplitude greater than x_2) is applied to the input of the automatic controller then it is easy to see (*fig. 27*) that at the output of the integrator there will be a trapezoidal voltage.

Fig. 28 shows some oscillograms recorded with a regulating circuit consisting of a process analogue, an automatic controller with three-step action and an integrator. Fig. 28*a* is the step-function response of the process analogue employed. A step disturbance at the input of the process analogue (corresponding to a sudden readjustment of the regulating unit) produces the variation x in the regulated quantity as illustrated in fig. 28*b*, and this in turn causes the voltage q at the input of the process analogue to change according to fig. 28*c*. The consequences of a step disturbance at the output of the process analogue are shown in figs 28*d* and *e*, respectively at the input and at the output of the process. In both these cases in the final position x comes to rest at a value between x_1 and x_2.

The examples that have been given here should be sufficient to show how valuable an aid the electro-analogue is for studying regulating processes. Details of the electric circuits of various parts such as the universal process analogue, the collecting stages, integrators and differentiators have purposely been omitted, since they will form the subject of a second article.

Summary. An electro-analogue is an apparatus with which electrical models of regulating devices can be built up and studied. The electro-analogue discussed here comprises two special models and one universal model of the process to be regulated, the latter model consisting of an electric network — mainly a delay network — the step-function response of which, displayed on the screen of an oscilloscope, can be given a shape identical to that of any process to be imitated. Further the electro-analogue comprises electrical models of automatic controllers with continuous action and of certain controllers with discontinuous action. In the model of the continuously acting controller a proportional term, a single and a double integrating term and a single and a double differentiating term can be realized.

After the model of the process has been given the right characteristic it is combined with one of the models of the automatic controller to form a closed regulating circuit. By applying a step disturbance at a suitable point it is possible to study with the aid of oscillograms the behaviour of the circuit at different values of the parameters. In particular it can be investigated under what conditions instability arises and how high the quality of the regulation is. Oscillograms are given for various cases, both with continuously and with discontinuously acting controllers.

A description of the electric circuits of the principal parts will be given in a subsequent article.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1934: W. de Groot: The radiation constants and the light equivalent of energy (Physica, The Hague **16**, 419-420, 1950, No. 4).

Hoffmann has drawn attention to the fact that the discrepancy between the experimental and the theoretical values of the Stephan-Boltzmann constant σ and the Planck constant c_2 would disappear simultaneously if the figure for the freezing point of gold (T_{Au}) were a few degrees higher than the internationally adopted value of 1336 °K. It is shown that the same holds true for the discrepancy between the experimental and the theoretical values of the light-equivalent of energy. At the same time it is shown why Van Dusen and Dahl obtained an experimental c_2-value in accordance with the theoretical value.

1935: C. J. Bouwkamp: On integrals occurring in the theory of diffractions of electromagnetic waves by a circular disk (Proc. Kon. Ned. Akad. Wetensch. Amsterdam **53**, 654-661, 1950, No. 5, Indigationes Mathematicae **12**, 199-206, 1950, No. 3).

Evaluation of integrals occurring in the rigorous theory of diffraction of a plane-polarized electromagnetic wave by a circular disk or by a circular hole in an infinite plane screen, if the radius of the disk or hole is small compared to the wavelength. Expansions are given in terms of Legendre functions, Bessel functions, and hypergeometric functions.

1936: C. J. Bouwkamp: On the theory of spheroidal wave functions of order zero (Proc. Kon. Ned. Akad. Wetensch. Amsterdam **53**, 931-944, 1950, No. 6; Indigationes Mathematicae **12**, 326-339, 1950, No. 3).

Spheroidal wave functions of order zero emerge from a separation of the wave equation in spheroidal coordinates in the case of axially symmetric wave problems. The author gives an account of the conventional theory of these functions. In addition, he draws attention to a new canonical system of spheroidal wave functions which has many advantages over the conventional systems.

1937: W. J. Oosterkamp: Dose measurements on contact therapy tubes (Acta Radiologica **33**, 491-506, 1950, No. 6).

The various causes of errors in contact-therapy dosimetry are analysed and methods for avoiding these errors are described. Extensive measurements have been carried out with the Philips apparatus in regard to: the absorption characteristic of the radiation in aluminium, the half-value layer for different added filters, the field distribution where the effect of a compensating filter is described, the determination of depth-dose data and of isodose curves in a water phantom, the influence of secondary radiation from the applicators and the dose rate of the stray-radiation around the tube caused by scattering from the patient. The dynamic electrometer used proved to be very useful for this kind of work. The results of the measurements are compared with those obtained by other authors.

1938: H. O. Huisman: Investigations into the vitamin A series, I. A new synthesis of vitamin A acid and vitamin A (Rec. Trav. chim. Pays-Bas **69**, 851-857, 1950, No. 6).

The chemical and physical properties of β-ionylidene acetaldehyde, the important key intermediate for the synthesis of vitamin A, are given. Starting from β-ionylidene acetaldehyde a new synthesis of vitamin A acid and vitamin A is described.

1939*: W. Elenbaas, J. Funke, Th. Hehenkamp, L. C. Kalff, A. A. Kruithof, J. L. Ouweltjes, L. M. C. Touw, D. Vermeulen, R. van der Veen: Fluorescentie-verlichting (edited by Prof. Dr. C. Zwikker), Meulenhoff, Amsterdam 1950 (262 pp., 173 fig.). (Fluorescent lighting, English edition in preparation.)

See Philips Techn. Rev. **12**, 369, 1950, No. 12.

1940: B. D. H. Tellegen: De krachten tussen twee stroomgeleiders (T. Ned. Radiog. **15**, 157-161, 1950, Nos 4 and 5). (The forces between two steady currents; in Dutch.)

The forces between two steady currents can be conceived of as the resultant of fictitious attracting forces and fictitious couples which the elements of current exert on one another; the attracting forces are proportional to the scalar product of the current elements divided by the square of their distance; the couples are proportional to the vector product of the current elements divided by their distance.

1941: J. H. van Santen and G. H. Jonker: Electrical conductivity of ferromagnetic compounds of manganese with perovskite structure (Physica, The Hague **16**, 599-600, 1950, Nos 7 and 8).

Report on measurements of electrical conductivity of polycrystalline ceramic compounds previously studied (see No. **1932** of these reports). Graphs are given showing $\log \varrho$ versus $1/T$ for $(La_{1-x}\ Sr_x)MnO_3$ with x as parameter and $\log \varrho$ versus x for $T = 100$ °K. A striking resemblance is to be observed between the curves showing the Curie temperature and saturation magnetisation versus composition. Similar sets of curves have been obtained for systems in which Sr is replaced by Ca or Ba.

1942: F. de Boer, J. H. van Santen and E. J. W. Verwey: The electrostatic contribution to the lattice energy of some ordered spinels (J. chem. Phys. **18**, 1032-1034, 1950, No. 8).

The electrostatic contribution to the lattice energy of some ordered spinels has been calculated. It is made plausible that also in the case of certain spinels that do not show any long-range ordering a considerable amount of short-range ordering is to be expected. As a consequence the lattice energy of "inversed" spinels containing different ions at octahedral positions is increased with respect to the corresponding "normal" spinels containing one type of ions at octahedral interstices.

1943*: E. J. W. Verwey: The role of the electric double layer in the behaviour of lyophobic colloids (Colloid chemistry — theoretical and applied, Vol. 7, Reinhold Publ. Corp., New York 1950; p. 47-66).

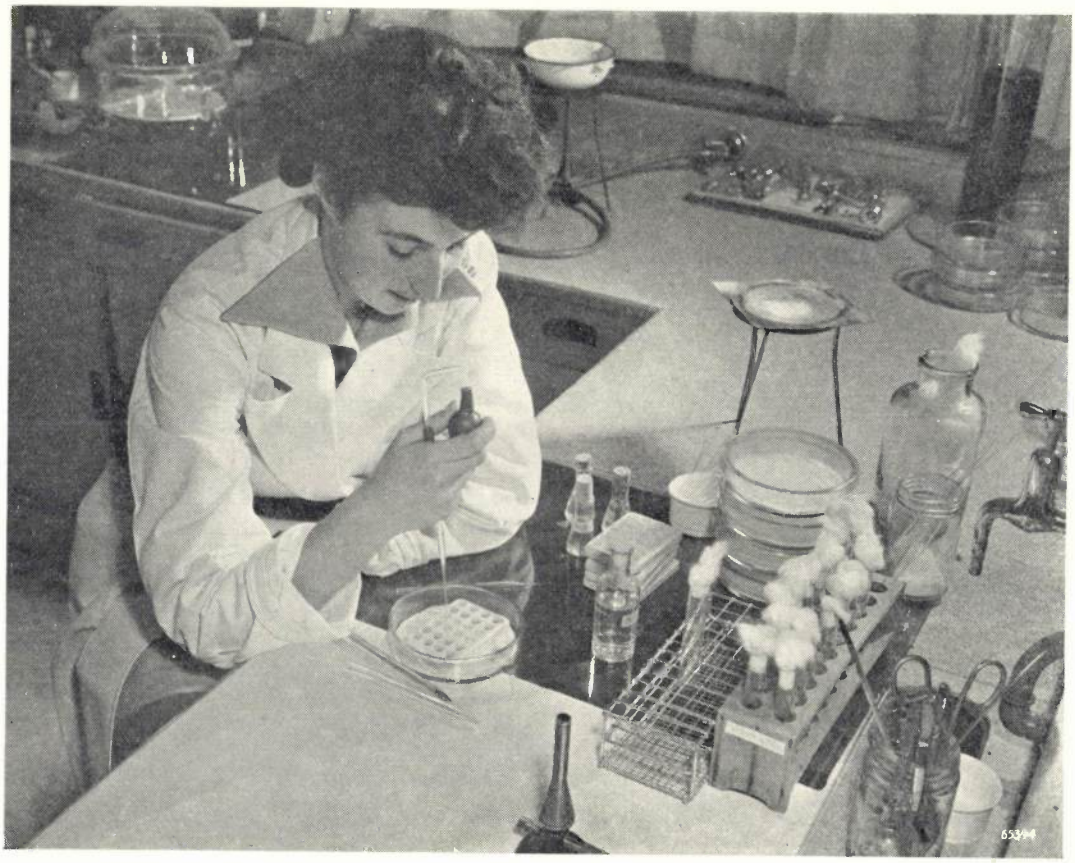
Discussion of the properties of the electric double layer of colloid particles and its influence on the interaction of particles (see Nos **1730**, **1769***, **1916** of these abstracts).

VOL. 12 No. 10, pp. 273-304 APRIL 1951

Philips Technical Review

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

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



THE MANUFACTURE OF VIRUS VACCINE AGAINST INFLUENZA

by A. J. KLEIN *) and E. HERTZBERGER *).

616.921.5-085.371

*For some time now a branch of the Philips Industries has been making Vitamin-D preparations **). During recent years this branch has developed into a many-sided industry turning out pharmaceutical products such as insulin, preparations of liver and hormones, plant-hormones, insecticides, weed-killers, etc. It is now engaged upon the production of virus vaccines, a peculiar form of mass production taking place partly in some large egg hatcheries and promising to be of great importance not only for the Netherlands but also far beyond its frontiers.*

Influenza as an epidemic

A number of infectious diseases occurring in the form of epidemics, such as cholera, plague, typhoid, smallpox, etc., have lost their grip on mankind thanks to counter-measures evolved by medical science. These counter-measures consist on the one hand in the prevention of infection by certain hygienic precautions and on the other hand

*) N.V. Philips-Roxane, Pharmaceutical-chemical Industry "Duphar", Weesp, Holland.

) See A. van Wijk, Lamp manufacture and vitamine research, Philips Techn. Rev. **3, 33-39, 1938.

in the immunization, by means of sera or vaccines, of individuals liable to infection.

During the last ten years it has been made possible to combat also influenza in this way. This disease often occurs in more or less harmless forms and as a result the public at large tends to underestimate its seriousness. The terrible epidemic of influenza that spread over the whole world in 1918 and 1919 and cost about 20 million people their lives is, however, evidence enough to show how important it is to have the means of combating this disease. Moreover, the seriousness of influenza is not to be judged solely from the number of deaths it causes. Influenza epidemics with low mortality figures, occurring, as they do, every other year or so, cause great losses to the national economy owing to the high percentage of workers laid up for weeks on end.

The possibility now offered of immunizing either certain individuals or whole groups of the people against influenza is therefore a great boon both from the medical as well as the social-economic point of view.

In this article something will be said about the principles underlying the preparation of the vaccine for immunization against influenza, the methods for the large-scale production of this vaccine and in particular those employed by Philips-Roxane at Weesp in Holland, and about the manner in which the vaccine is to be used in the event of an epidemic.

Immunization; serum and vaccine

What are the principles underlying immunization against an infectious disease?

The disease is caused by a morbific germ, in the form of a bacillum or a virus, getting into the body and very rapidly multiplying itself. The body reacts to this by producing specific antibodies rendering the morbific germs harmless. Roughly speaking the process of the disease can then be regarded as a contest between the morbific germs and the antibodies: if the former multiply themselves at a faster rate than the latter are produced then they win, whilst in the other case the patient wins and recovers. If a person is already carrying a quantity of the specific antibodies before the morbific germ gets into his body, then as soon as there is an infection that person has a good start, so to speak, and better chances of throwing off the attack — he is more or less "immune" to the disease.

One can become immune in three ways:

(a) Once a person has had an attack of the disease and fully recovered from it there is usually a surplus of antibodies left in the body and this surplus continues to be active for a time, so that one is naturally immune against another attack for some length of time and possibly, in the case of some diseases, for the rest of one's life.

(*b*) A person can be injected with a dose of antibodies taken from some other persons or from test animals who have passed through the process of the disease; a preparation of antibodies sufficiently concentrated for this purpose can usually be extracted from the blood of a test animal, and this supplies the preparations known as "sera" or anti-sera.

(*c*) The body of the person to be immunized can be incited to prepare antibodies without rendering that body ill. This is done by injecting a "vaccine", i.e. a preparation of inactivated germs. It is indeed surprising that this should be possible, considering that the morbific germ is "killed" and can no longer multiply itself, but it is nevertheless still capable of bringing about the production of the specific antibodies. In fact it is not so surprising when it is borne in mind that there are also "dead" substances (proteins) capable of causing the formation of corresponding "antibodies" in the human body. Anyhow this method is probably the most important weapon available to the physician in combating epidemics, more so than the serum method. Since it has not yet been found possible to extract antibodies in the pure state for any disease, and only for a few diseases is it possible to extract them in strong concentrations, usually rather large quantities of a serum (some tens of cm^3) have to be injected and owing to the sensitivity of the human body for foreign proteins this not infrequently gives rise to complications. For vaccination, on the other hand, usually only a few drops of the vaccine are sufficient because of the high concentration in which it can be produced, thanks to the multiplying power of the morbific germs (not yet inactivated!). Furthermore, antibodies taken from other people, and especially when they are taken from animals, being "generic-foreign material", are fairly soon expelled from the human system, so that the protection afforded by a serum (passive immunity) is of short duration — at most six weeks — whereas "active" immunity obtained by vaccination may last many months and even years.

The influenza virus

It is only possible to prepare a vaccine (and also a serum) after the germ causing the disease concerned has been identified, or rather — in the case

of the morbific germ being unknown — after it has been possible to isolate it. Such was first possible in the case of influenza in 1933, when Smith, Adrewes and Laidlaw [1]) succeeded in extracting in a more or less pure form a virus from the liquid used by influenza patients for gargling. Test animals infected with this virus "caught" influenza. While the patients were recovering these investigators also succeeded in detecting in their blood antibodies against the virus found; test animals injected with virus preparations previously mixed with the blood extract did not become ill.

Further investigations carried out by a number of research institutions in America, Great Britain, the Netherlands and other countries proved that in the case of influenza it is not a matter of one morbific germ but that there are several different "strains" of virus, distinguished, inter alia, by the fact that they cause the formation of different antibodies. According to their degree of relationship these strains of virus can be classified under a few groups which have been named influenza-A-virus, influenza-A'-virus and influenza-B-virus. A virus can be neutralized not only with its own antibody but also with antibodies of other viruses belonging to the same group. The practical importance of this fact will be explained later. Suffice it to say here that a number of laboratories now have preparations of the isolated strains of virus at their disposal, from which by cultivation they are able to produce the quantities of virus required for experiments and for the manufacture of vaccines and anti-sera.

Fundamentals of the preparation of virus vaccine

In essence the cultivation of a virus consists in causing it to multiply itself in a suitable medium. With the object of vaccination, this cultivation is necessary for two reasons: firstly, the extraction, purification and concentration of a virus from the excretions of influenza patients is a cumbersome process and yields only comparatively small quantities, whilst secondly it has been proved that for an effective immunization highly concentrated vaccines have to be used. Prior to 1941 such highly concentrated vaccines were not available and this is the reason why the first attempts to immunize groups of the population met with little success, although with mice and ferrets — the test animals mostly used for this work — a decided immunizing action of the inactivated virus preparations had been observed immediately.

[1]) W. Smith, C. H. Andrewes and P. P. Laidlaw, Lancet **225**, 66, 1933.

The cultivation of a virus on any large scale was a problem never encountered before. Good methods were known for cultivating all sorts of bacteria, e.g. in bouillon or blood which may be made into a solid culture medium with agar-agar, at certain temperatures. One of the essential differences between a bacillum and a virus, however, is that the latter can only multiply on or in living tissues. The customary simple culture media for bacteria were therefore unsuitable and use had to be made of test animals or of tissues derived from test animals and kept alive in a suitable medium (mostly of rather complicated composition, i.e. the so-called tissue cultures). For a time the virus was cultivated, for instance, in the lungs of mice. This technique was far too cumbersome for the large-scale production of vaccine, and the purification (removal of undesirable proteins) and sterile preservation of the product (exclusion of foreign germs which would take part in the multiplication and make the vaccine useless) were difficult problems. Consequently the result could not be called satisfactory.

An excellent idea, which eliminated at once most of the difficulties, was to use embryonated chicken eggs as medium for the cultivation [2]). Here we have indeed a living tissue, easily obtainable at all times in large quantities, contained in a shell guaranteeing sterility — veritably the egg of Columbus. When this idea came to be further investigated it appeared that with a suitable choice of the spot inside the egg also an excellent opportunity was afforded for the purifying of the virus solution obtained. Since 1940 very effective methods have been developed upon this principle for the cultivation of influenza virus. It is such a method — worked out for the greater part empirically — that is being employed in the mass manufacture of influenza virus vaccine by Philips-Roxane at Weesp (Holland).

Before proceeding to describe the details of the method followed, it is to be pointed out that the cultivation of the virus is only one of the steps in the preparation of the virus vaccine. Other essential steps are: the purifying of the virus, its inactivation, possibly drying, and last but not least the assay and dilution. These processes, too, will be briefly described.

Incidentally, it may be mentioned that once the technique of manufacturing a virus vaccine has been mastered the possibility of immunization also against other diseases caused by a virus comes within the reach of the medical profession. A vaccine

[2]) W. Smith, Brit. J. Exp. Path. **16**, 508, 1935.

can be made against parotitis (mumps) for instance, the methods for the preparation of which are entirely analogous to those followed for influenza vaccine.

Manufacture of the vaccine

Cultivation of the virus in embryonated chicken eggs

Fig. 1 is a cross section of a fertilized chicken egg after ten days' incubation (hatching normally takes 21 days). The most favourable part of the egg for the development and subsequent purification of the virus is the membrane of the allantoic cavity. This cavity, filled with the clear allantoic fluid, almost entirely envelops the embryo and takes up all the waste products while the embryo is developing, thus acting, as it were, as its bladder. On the tenth day of incubation the virus is injected into this cavity, where in two days' time it multiplies about 100,000 fold. In this way, from one or two eggs (according to the strain of the virus) an amount of virus is obtained sufficient for immunizing one person.

An enormous number of eggs are therefore required for manufacturing vaccine on a scale calculated for the immunization of a whole population. The incubation of the eggs for the required period of ten days and the cultivation of the virus during

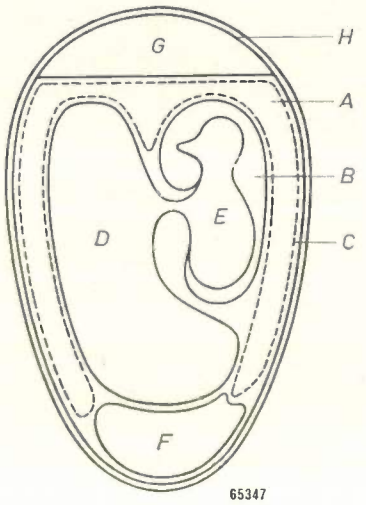


Fig. 1. Diagrammatic cross section of a fertilized chicken egg after 10 days' incubation. *E* embryo, *D* yolk sac, *B* amnion cavity, *A* allantoic cavity, *C* allantoic membrane, *F* albumin, *G* air space cell, *H* shell membrane.

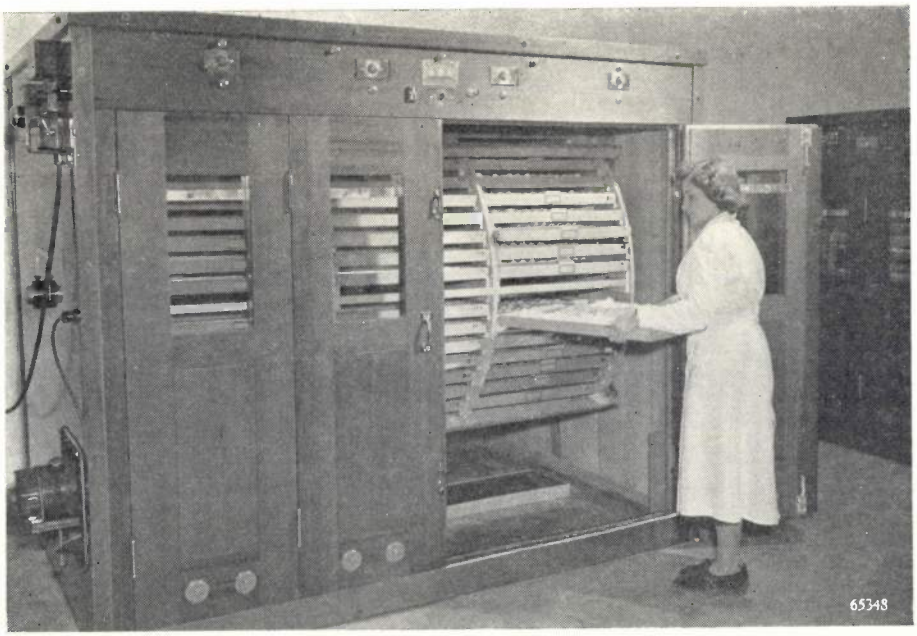


Fig. 2. Machine for the simultaneous hatching of 16,000 eggs. The temperature inside the cabinet is maintained at a level of 37.5 °C. At the bottom are trays of water providing for the necessary humidity of the air, so that the eggs do not lose any of their liquid through the pores of the shell. The trays filled with eggs are placed in a large drum which can be rotated through an angle of 90 degrees: in one position of the drum the eggs are tilted 45° to one side and in the other position 45° to the other side. This turning of the eggs (in imitation of what the brooding hen regularly does) has to be done every day to prevent the heaviest parts of the egg from settling at the bottom of the shell. A wooden framework rotating around the drum provides for a uniform distribution of the heat and moisture content of the air inside the cabinet.

two days is therefore done in large incubators such as are used in the poultry-raising industry, each having a capacity of, say, 16,000 eggs (see *fig. 2*).

After three days' of incubation all eggs are examined in a darkened room, one by one, in a strong beam of light from a mercury lamp (see *fig. 3*). *Fig. 4* is a picture of a fertilized egg as examined in this way; the blood vessels formed during the first three days of incubation are clearly seen. At this stage it is thus possible to see whether an egg has been fertilized or not, and those that have not been fertilized are rejected. After ten days' incubation the eggs are examined again and marked at the spots where "inoculation" and "harvesting" is to be done. This is necessary because the eggs are inoculated by drilling a small hole in the shell and then injecting a few drops of a virus preparation: when this is being done care has to be taken not to puncture one of the blood vessels. The laboratory assistant examining the egg therefore marks it with a cross at a point suitable for the injection at the right height and at a safe distance from the blood vessels. Before injection takes place a second small hole has to be made in the air space of the egg (see fig. 1), so that as the virus is injected a corresponding amount of air can escape. For the subsequent harvesting the boundary of the air space is also marked out on the shell at this second examination of the egg.

The drilling of the holes is illustrated in *fig. 5* and the process of injection in *fig. 6*. Injecting is done in a glass-top case with two openings through which the assistant can pass her arms. After the inoculation the two holes in the egg are sealed with paraffin wax. Inside the case is a germicidal lamp (the TUV lamp), the ultra-violet rays from which keep the air and everything else inside the case sterile [3]); it is of the utmost importance that no living bacteria or active undesired viruses on the egg shell or on the point of the injection needle should get into the egg and there multiply together with the influenza virus (of course at this stage the influenza virus itself must not come under the germicidal action of the lamp rays). To keep the air inside the case sterile in spite of the arm holes, a certain pressure is maintained in the case by means of a ventilator; the stream of air entering at the back passes through a filter, where all dust is filtered out, past a second germicidal lamp and thus in a perfectly sterile condition into the work compartment, from which it flows out through the arm holes (see *fig. 7*).

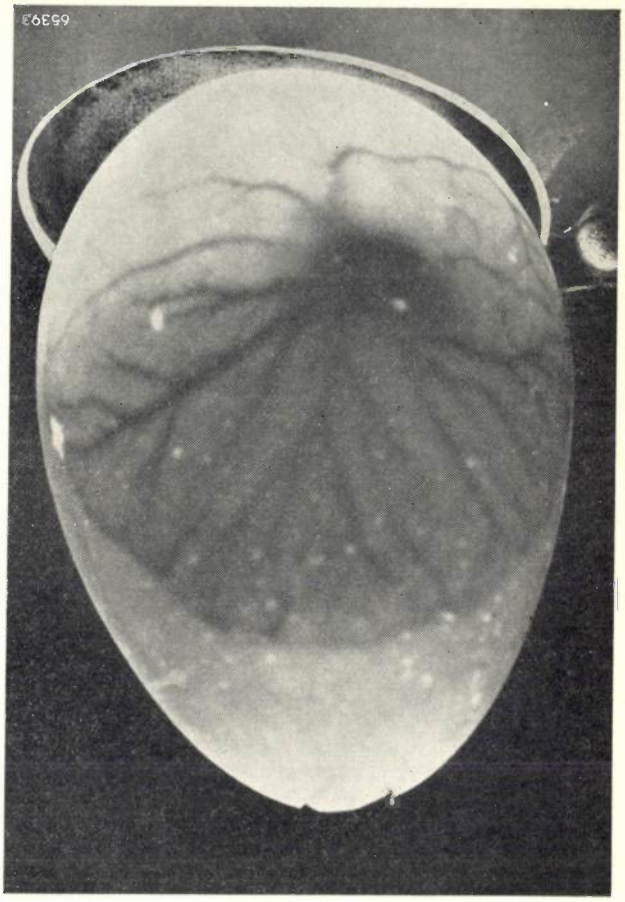


Fig. 4. An egg as seen when examining it after three days' incubation. The blood vessels are clearly seen. The boundary of the air space cell cannot be seen on the photograph owing to the halo effect.

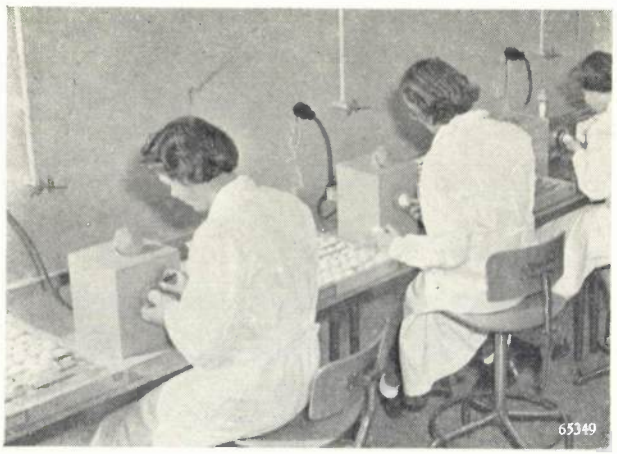


Fig. 3. Examination of the eggs in a strong beam of light to see whether they have indeed been fertilized; the same devices are also used for marking the eggs for subsequent inoculation and harvesting.

[3]) For a description of these lamps and their operation see J. Voogd and J. Daams, Inactivation of bacteria by ultraviolet radiation, Philips Techn. Rev. 12, 111-119, 1950 (No. 4).

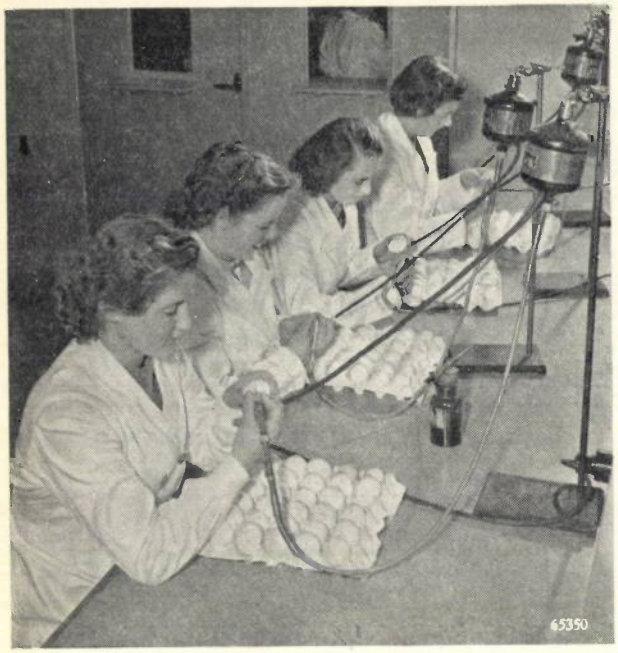


Fig. 5. The necessary holes are drilled in the eggs with a kind of dentist's drill.

There is not much to be said about the actual process of cultivation, which takes place in two days after the injection of the virus into the egg. It is rather peculiar that the best results are obtained when the incubating temperature during the cultivation period is kept at 36.5 °C, whilst the normal incubating temperature, maintained during the first 10 days of incubation, is 37.5 °C.

After the two days' cultivation the eggs are kept 24 hours in a refrigerator, so that the embryo dies off and the rest of the contents of the egg "sets" more or less, this being desired for the "harvesting". For the harvesting the part of the egg shell covering the air space is cut off without damaging the contents of the egg. Then with a pipette, connected via a flask to a water-jet pump, the liquid contents of the allantoic cavity are drawn off into the flask. This, too, is done in a sterilized case with arm holes, at which a number of laboratory assistants work together (see *fig. 8*). One flask can contain the culture from about 300 eggs.

Fig. 6. The case, kept sterile with TUV lamps, in which the incubated eggs are inoculated with the virus to be cultivated. The glass top does not allow the rays of the 2537 Å wavelength to pass through. The arm holes can be moved a little to the side to provide a comfortable position for the assistant's arms.

After sterilization by heating, the eggs from which the allantoic fluid has been extracted can be worked up into cattle food.

The virus preparations required for the inoculation of the eggs are kept in a refrigerator and by cultivation, likewise on incubated eggs, in a small thermostat (miniature incubator) can be constantly renewed.

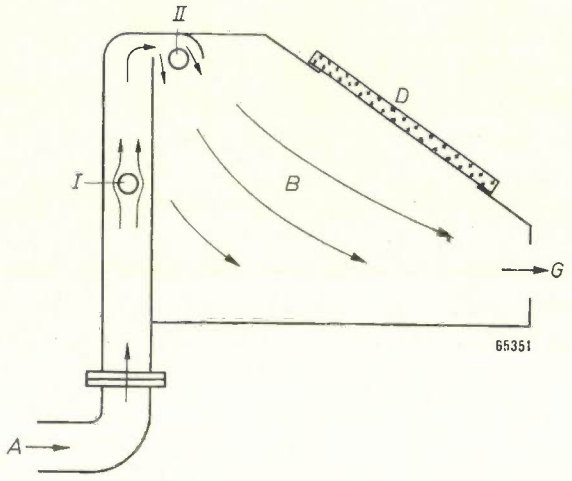


Fig. 7. Cross-sectional diagram of the inoculating case. The work compartment *B*, with glass top *D* and arm holes *G*, is kept sterile by the TUV lamp *II*. A certain pressure is maintained inside the case by the admission of filtered air at *A* with the aid of a ventilator. This air is sterilized as it passes the TUV lamps *I* and *II* and leaves the case through the arm holes.

Purifying and assaying the virus preparation

For purifying the virus, that is to say the removal of most of the undesired proteins present in the solution that has been collected, a process is applied which might be called "differential centrifugation". In a fairly large, normal, centrifuge with a speed of 3000 r.p.m. the flocculent precipitates from the virus solution are deposited on the wall of the machine. The remaining liquid containing the virus is run through a small centrifuge rotating at a very high speed (Sharples super-centrifuge, 50,000 r.p.m.); see *fig. 9*. This centrifuge is highly selective, all the molecules over a certain weight — and those only — being deposited on the wall. Thus the virus (which in this connection can be regarded as a heavy protein molecule) is deposited on the wall of the rotor in an extremely pure state, being subsequently shaken off with a suitable liquid. The speed of the centrifuge and the rate of flow have to be very precisely adjusted and continually checked,

Fig. 8. Case, sterilized with TUV lamps, in which the cultivated virus is extracted from the eggs. Each of the assistants working at the case uses a pipette, connected to a suction pump, with which the liquid is drawn out of the allantoic cavity of the "decapitated" egg into the flask standing in the middle of the case. This flask can hold the culture taken from 300 eggs. In each suction line is a small intermediate flask which has to be inverted after every fourth egg for its contents to be transferred to the big flask. If one of the eggs happens to be a failure (for instance through a blood vessel bursting) the intermediate flask is replaced by a clean one, so that the contents of the big flask are not spoiled.

so as to ensure the right sedimentation. Since, during the flow of the liquid through the centrifuge, outside air is apt to be drawn in (see fig. 9*b*), also the centrifuge is set up inside a case sterilized with the aid of TUV lamps (*fig. 10*).

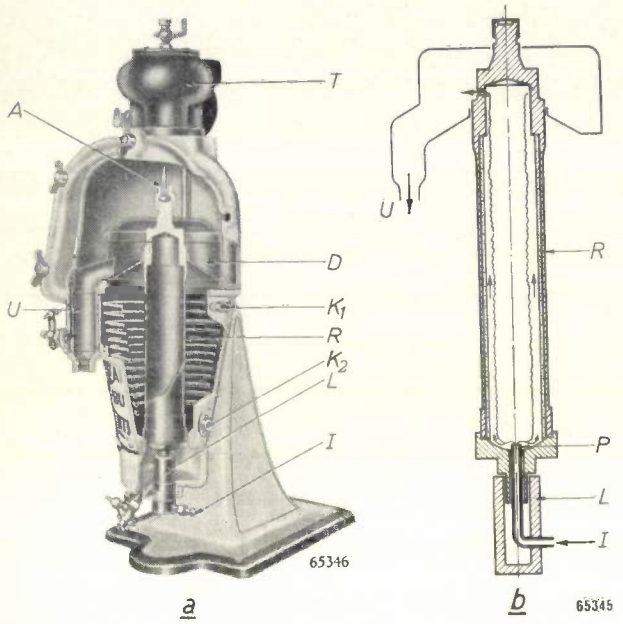


Fig. 9. *a*) The Sharples supercentrifuge. *T* driving steam or compressed-air turbine, *A* shaft rotating at 50,000 r.p.m., *R* rotor mounted on the shaft and through which the liquid to be centrifuged flows, *I* inlet, *U* outlet, *D* cover, *L* bearing leaving sufficient play for the rotor to find its own ideal axis of rotation, K_1 and K_2 water inlet and outlet for the cooling coil. In *b*) the sedimentation in the rotor (drawn very roughly) is made a little clearer. The liquid entering the rotor at *P* follows the rotation of the rotor almost immediately.

The highly concentrated virus preparation thus obtained is shaken with a small addition of formaldehyde, thereby completely inactivating the virus. After adding a preservative (sodium ethylmercuri-thiosalicylate) the virus solution can be kept in a refrigerator until it is needed.

Another method of preparing the virus for storage is by drying. Of course the water cannot be expelled from the solution by heating because then the virus is damaged. Boiling under reduced pressure is not practicable either, because then too much virus is lost (owing to excessive heating, concentration of salts or other causes). Very good results are obtained, however, by freezing the virus solution (in a vessel cooled with carbon dioxide) and then evaporating the ice by connecting the vessel to a vacuum pump. The dry residue containing the virus is then kept in a refrigerator at a temperature of —20 °C.

The last, and very important, step in the manufacture of the vaccine is the determination of the activity of the suspension (assaying) and the dilution to the desired strength.

There is a simple method of determining the concentration of an influenza-virus suspension, namely the Hirst test [4]). This is based upon the discovery that influenza-virus suspensions cause the red blood corpuscles of chickens to clump. This phenomenon allows of the quantitative measuring of the concentration, and the Hirst test has therefore become an indispensable means of continually checking the manufacture of the vaccine and of making the desired dilutions (see frontispiece). This cannot, however, be said to be an assaying in the strict sense of the word: virus suspensions of the same concentration but from different cultures may vary considerably as far as their activity as vaccine is concerned, since the virus may have a different degree of antigenic action, i.e. a different power (retained or originally possessed) to produce antibodies. It is therefore necessary to determine the activity of the virus suspension as vaccine by means of test animals.

This is done according to the method prescribed by the National Institute of Health in the U.S.A., which has also been made compulsory by the State Institute of Public Health in the Netherlands. A number of mice are inoculated with different doses of the vaccine and after 14 days samples of their blood are taken. This blood serum is mixed with active influenza virus and dripped into the nostrils of non-inoculated mice. The serum from the vaccinated mice should neutralize the active virus, in which case the mice to which the mixture has been administered will stay alive. The dose of vaccine just neutralizing the virus is taken as the standard strength on which further dilutions are to be based.

This assaying is done also in the laboratories at Weesp. To conclude the preparation of the vaccine the prescribed tests are carried out — partly on animals — for making sure that the vaccine does not contain any living virus, any living germs of some other genus, or any other harmful substances of whatever nature. Not until then can the vaccine be released for medical use.

This is not the place to go into the manner in which the vaccine is injected into human beings. All that need be said here is that the persons to be vaccinated must not be oversensitive to chicken protein. Fortunately this is only the case with about 0.5% of the population, and any over-

[4]) So long as a virus cannot be completely isolated it is of course not possible to determine its quantity by weighing, so that one has to have recourse to such means as biological effects, specific absorption of light or fluorescence, or suchlike.

sensitiveness (allergic reaction) can be determined by first injecting a very small dose into the skin (the actual vaccination is by subcutaneous injection).

Combating influenza with the virus vaccine

The fact that in the case of influenza it is not a matter of just one single virus but of various strains is of great influence in the combating of the disease. It cannot be known in advance what strain of the virus will be responsible for a future epidemic.

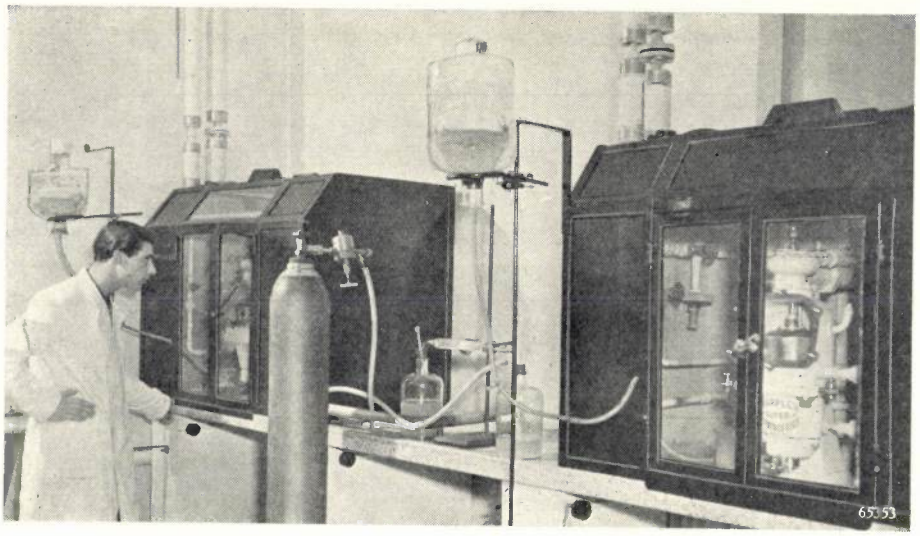


Fig. 10. Cases, kept sterile TUV lamps, in which the Sharples supercentrifuges are mounted. The centrifuges have to be easily accessible because at regular intervals the speed has to be checked and if necessary corrected.

Therefore, to make an influenza vaccine which can be held in readiness to cope with the first attack of an approaching epidemic or, in normal times while there are still no signs of an epidemic, to safeguard against possible future infection people for whom influenza involves particular risks, this vaccine has to be "polyvalent", that is to say it has to contain various suitably chosen strains of virus which together have the widest possible antigenic action.

To hold in readiness this polyvalent vaccine is not, however, an ideal solution. It is true that in the event of an epidemic breaking out the majority of the people vaccinated will be safeguarded against the infection if they have been vaccinated early enough and if the strain of virus causing the epidemic is susceptible for one of the antibodies created. But if the polyvalent vaccine is made up, for instance, from a mixture of three strains then every person has to be inoculated with three times the amount of vaccine that would be needed in the case of a "monovalent" vaccine. This is to be deprecated on account of the undesired secondary reactions, which in any case have to be limited to a minimum, and the more so because with a given capacity of the factory producing the vaccine only one third of the number of people can be immunized which otherwise could be treated with a monovalent vaccine.

Furthermore, it may happen that the epidemic is caused by a strain of virus as yet entirely unknown, a mutant such as seems to appear fairly regularly every ten years and which may not react at all to the antibodies formed by the polyvalent vaccine. In that case everyone who has already been vaccinated has to be revaccinated.

With a view to this situation, in 1948 the World Health Organisation established a service for reporting to all countries any outbreak of influenza in any part of the world and for taking immediate steps to isolate and identify the virus. Thus, in all countries to which the epidemic may be expected to spread in the course of some time it is now possible to proceed at once to the production of a monovalent vaccine against the identified morbific germ of the disease.

This means that the industry making the influenza vaccine is obliged to equip itself for coping with very heavy "peak loads". In order to meet this requirement without involving the installation of large permanent plant that would be standing idle most of the time, Philips-Roxane have estab-

lished a remarkable organisation. The manufacture of the virus vaccine depends for a large part on the use of large incubators for hatching the thousands of eggs needed. Such incubators, however, are already available in the poultry industry, where as a rule they are idle for several months of the year, just in the summer and autumn when epidemics of influenza commonly break out. It has therefore been arranged with several hatcheries in the country (from whom the fertilized eggs for cultivating the vaccine are procured) for their machines to be used in the vaccine production process in case of need. By temporarily increasing the number of assistants making up the inoculating and harvesting teams (the inoculating teams work in the hatchery, whilst for harvesting the eggs are transported to the factory) the production capacity can now easily be raised to a multiple of the normal.

Even with such a greatly increased capacity it is still impossible, within the short time available before the reported epidemic spreads to the country, to produce sufficient vaccine for immunizing the whole population. But then, after all, this is not really necessary. Much harm can already be avoided by immunizing those groups of people most exposed to the infection and certain individuals whose incapacitation would help the epidemic to spread and thus aggravate the general dislocation of public services and industry. The vaccine available will therefore be reserved in the first place for the vaccination of doctors, nurses and suchlike, for personnel of electricity works, telephone exchanges and the foodstuffs industry, further for people living or working in communities, such as the military forces, seamen, pupils in boarding schools, workers in large factories, etc. For the same groups of people a preventive vaccination in normal times might also be considered. This would have to take place every year, because between 6 and 12 months after the injection immunization begins to decline.

Summary. As is the case with other infectious diseases, immunization is now also possible against influenza. The vaccine, a concentrated suspension of the inactivated influenza virus, is being manufactured on a large scale by cultivating the virus in the allantoic cavity of embryonated chickens' eggs. The injection of the virus into the egg, the cultivation and the harvesting of the greatly multiplied virus, as also the purification of the virus suspension and the assaying for correctly dosing as vaccine are described in this article. Since there are a number of strains of the influenza virus, in normal times a polyvalent vaccine is made containing a mixture of different strains and thus offering immunization against various strains of the virus. With the approach of an epidemic of influenza the specific strains of virus responsible for it is reported by an international organisation to the threatened countries, where steps can immediately be taken for producing a monovalent vaccine against that particular strain and vaccinating the population as far as necessary and possible.

THE ELECTRICAL RECORDING OF DIAGRAMS WITH A CALIBRATED SYSTEM OF COORDINATES

by B. G. DAMMERS, P. D. van der KNAAP and A. G. W. UITJENS.

621.317.755: 621.3.012

The construction of diagrams from a series of measurements demands the closest attention and takes a considerable time. Electronic engineering in recent years has enabled the construction of calculating machines which can take over some of the work of the human mind, and it also offers the possibility of constructing an apparatus which can very quickly take measurements and record their results in the form of a diagram.

Several articles have already appeared in this journal describing apparatus by means of which diagrams or graphs can be traced on the screen of a cathode-ray tube [1][2]).

The tracing of such diagrams is done by periodically varying the "independent variable", measured by means of a proportional voltage, from the minimum to the maximum value. The "dependent variable" (also measured electrically) then likewise varies periodically through a certain range. After amplification these two voltages are applied to the horizontal and vertical deflection plates of a cathode-ray tube, thereby producing a curve on the screen.

In the apparatus described here, a calibrated system of coordinates in the form of a lattice is produced on the screen simultaneously with the tracing of the diagrams. This allows of an accurate numerical interpretation of the curves, which with the apparatus hitherto employed was more difficult and less accurate, owing to the lack of calibrating points.

Although this diagram-tracing apparatus can also be used for many other variables that can be measured electrically, this article will only deal with its application for recording the characteristics of electronic valves, where the quantities to be measured are already voltages or currents. This apparatus has proved to be of great value in the electronic-valve development department, where new types and modified forms of valves are frequently being examined.

Compared with the usual method of constructing a diagram point by point from meter readings, the use of the cathode-ray tube for this purpose has the advantage that a curve can be plotted very quickly. This is highly important when, in the case of static measurements, the valve is likely to be overloaded and consequently damaged.

As was the case with the apparatus previously described, a family of I_a-V_a curves is simultaneously recorded for different values of the grid voltage of the valve. For this purpose, with the apparatus referred to in footnote [1]), the grid voltage was rapidly changed in steps with the aid of a rotating commutator. Here this mechanical method of switching has been replaced by an electronic system, in view of the higher switching frequency required for the simultaneous tracing of the coordinate lattice. In place of the "impulse tube" employed by Douma and Zijlstra[2]), normal amplifying valves are used.

Principle

The main parts comprising the diagram-tracing apparatus are indicated in the block diagram of *fig. 1*. The anode of the valve T under test receives from an alternating-voltage generator G, and a direct-voltage source V_a, a periodically changing positive voltage. This voltage is also applied to the horizontal deflection plates of the cathode-ray tube K via a deflection amplifier A_H. The voltage across the resistor R in the anode circuit is applied to the vertical deflection plates of K, also via an amplifier (A_V). Thus the coordinates of the spot describing a curve on the screen due to the changing anode voltage and anode current are at any moment proportional to the instantaneous values of these two electrical quantities.

The circuit also contains direct-voltage sources V_{g2} and V_{g3}, for applying the correct voltages to the screen grid and the third grid of the valve, and a step-voltage generator V_g supplying to the control grid a series of different direct

[1]) H. van Suchtelen, Applications of cathode ray tubes IV, Philips Techn. Rev. 3, 339-342, 1938.
A. J. Heins van der Ven, Testing output valves, Philips Techn. Rev. 5, 61-68, 1940.

[2]) Tj. Douma and P. Zijlstra, Recording the characteristics of transmitting valves, Philips Techn. Rev. 4, 56-60, 1939.

voltages, so that a family of I_a-V_a curves for different values of the parameter V_g are traced at short intervals.

The tracing of the set of curves is alternated by the tracing of the lines of the coordinate lattice. This is made possible by connecting at the point b_2 in the leads to the deflection amplifiers A_H and A_V electronic switches (not shown in fig. 1), which will be discussed later.

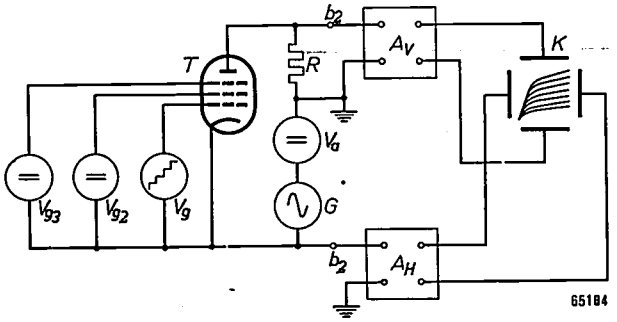


Fig. 1. Block diagram of the apparatus for measuring characteristics. T valve under test; V_a rectifier for supplying the anode direct voltage; G alternating-voltage generator for scanning the I_a-V_a-characteristic; R resistor across which a voltage is developed proportional to the anode current of the valve T; K cathode-ray tube; A_V amplifier for the vertical deflection of the beam; A_H amplifier for the horizontal deflection; V_g step-voltage generator for supplying the control-grid voltage of T; V_{g2}, V_{g3} rectifiers for the auxiliary grid voltages of T.

The tracing of the lines of the lattice is done in a similar way to the tracing of the curves, with this difference that one of the two deflection voltages is kept constant while the other varies periodically, in this case sinusoidally.

Twenty-five complete pictures (curves with lattice) are displayed per second. Each picture period is divided into four parts of 1/100 second, called quarters. In the first quarter the vertical lines of the lattice are traced, in the third quarter the horizontal lines, and in the second and fourth quarters the curves that are to be measured.

The three pairs of signals, viz. for the horizontal and vertical deflection (a) of the characteristics, (b) of the vertical lines of the lattice, and (c) of the horizontal lines, are conducted by three pairs of channels to the previously mentioned two electronic switches, which pass the pairs of signals in turn to the two deflection amplifiers of the cathode-ray tube.

The direct-voltage components of these three signals take part in the amplification up to the output of the electronic switch, thus making it possible to compare the direct-voltage components of the three pairs of signals: when there is any inequality in the switched voltages then, upon switching over, a voltage surge occurs at the output of the switch, and this causes a discontinuity in the curve traced on the screen; where there is no such discontinuity this indicates that the switched voltages are equal.

This possibility is used for calibrating the lattice of coordinates, by comparing the direct voltage required for each line of the lattice with an accurately known, constant, direct voltage, as will be explained when describing the calibration.

Beyond the switches only the alternating-voltage components are amplified by the two deflection amplifiers, so that the centre of the picture is automatically brought into the centre of the screen of the cathode-ray tube.

As already observed, switching is done by means of two identical electronic switches, H and V, periodically passing the necessary signals to the horizontal and vertical deflection plates (see *fig. 2*). These switches each have three channels with picture-signal inputs b_1, b_2 and b_3 and three switching-signal inputs s_1, s_2 and s_3. The three channels for the picture signals are periodically opened and closed by square-wave pulses, supplied by the switching-signal generator G_1, being applied to the switching-signal inputs. The picture signals for the lattice are supplied by a step-voltage generator G_2 and an alternating-voltage generator G_3 with a fairly high frequency (7020 c/s). The form of the pulses produced by the three generators is shown in fig. 2. The switching-signal generator G_1 supplies negative pulses for opening the picture channels. In the first quarter the channels *1* are opened, the plates for the horizontal deflection receiving a step-voltage from G_2 and those for

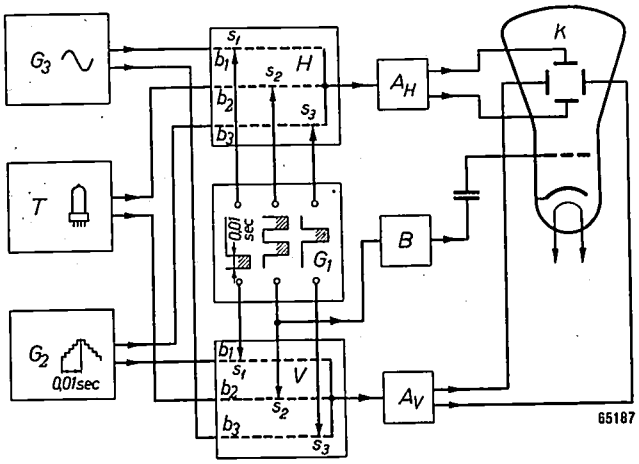


Fig. 2. Basic principle of the diagram tracer. H electronic switch for the horizontal deflection; V ditto for vertical deflection; G_1 switching-signal generator; G_2 step-voltage generator; G_3 7020 c/s generator; A_H amplifier for the horizontal deflection; A_V amplifier for the vertical deflection; B intensity control; K cathode-ray tube; T valve under test.

the vertical deflection a sinusoidal voltage from G_3. A series of vertical lines are then traced on the screen of the cathode-ray tube.

In the second and fourth quarters the channels *2* are opened and the curves are traced. In the third quarter the channels *3* are opened, the plates for the horizontal deflection receiving a sinusoidal voltage and those for the vertical deflection a step voltage, thereby tracing a series of horizontal lines on the screen.

The variation of the voltages on the two pairs of deflection plates is shown diagrammatically in *fig. 3*, where it is assumed that only one curve is being traced.

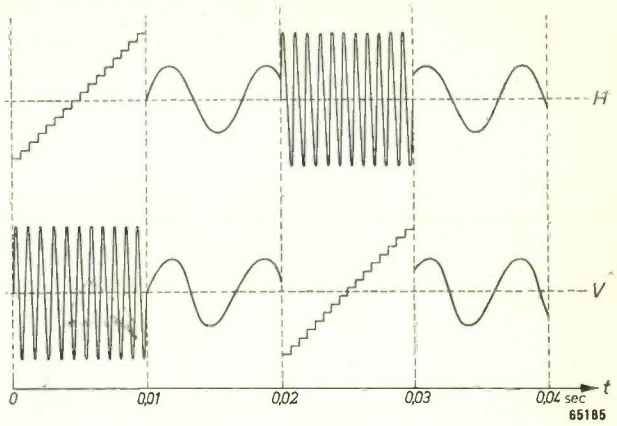


Fig. 3. Waveform of the deflection voltages when tracing one curve with a lattice of coordinates. *H* voltage on the horizontal deflection plate; *V* ditto for the vertical deflection.

As we have already seen, for the tracing of the lattice, step voltages are used. In the case where 13 lines (12 intervals) are traced the frequency in the succession of the steps is $13 \times 50 = 650$ c/s. The amplifiers have to amplify a considerable number of harmonics of this frequency with equal gain; the frequency band of the amplifiers is, of course, limited in width, so that it is advisable to keep the frequency of the steps as low as possible, thus to choose the smallest possible number of pictures per second. The lower limit of the picture frequency is determined by the persistency of

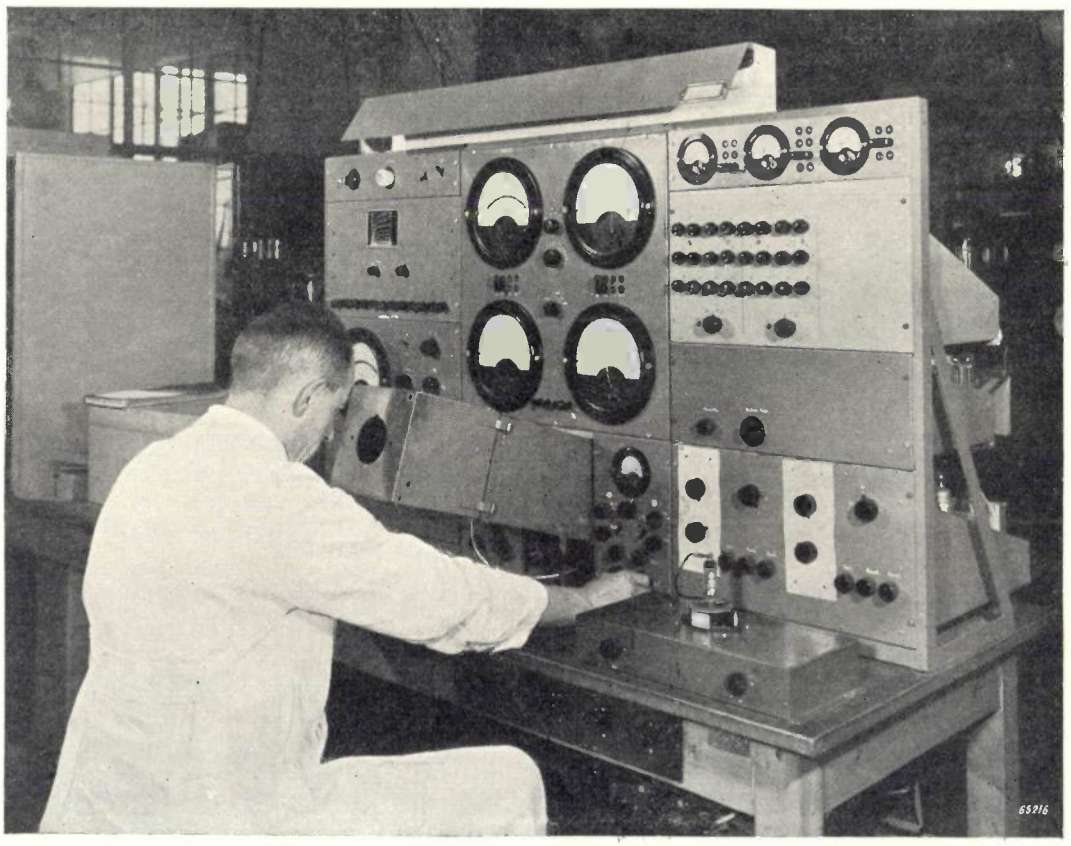


Fig. 4. Measuring and control panels of the diagram tracer. In the middle at the bottom the metal case containing the cathode-ray tube, type DW 16-2. On the bench on the right the valve under test. At the top on the middle panel the meters recording the filament current and the screen-grid voltage of the valve under test. On the right-hand panel at the top control meters for the screen-grid current, anode current and anode voltage, and underneath these the controls for adjusting the steps of the step-voltage generator for the lattice. On the left-hand panel at about the same level are the controls for adjusting the steps of the step-voltage generator supplying the grid voltage for the valve under test.

the screen and of the visual impression and also by the linearity in amplitude and phase of the amplifiers and electronic switches at low frequencies; 25 c/s has been found to be a good compromise.

In the block diagram (fig. 2) an "intensity control" B is indicated. This allows of the brightness of the lattice being varied with respect to that of the family of curves. If no special measures were taken, the curves — when there are only a few in the series — would be much brighter than the lattice and as a consequence the photographic recording of the picture on the screen would be over-exposed. To avoid this a square-wave control signal is applied via an amplifier to the grid of the cathode-ray tube, by means of which the intensity of the beam is reduced during the second and fourth quarters. This control signal has the same form as the signal that is used for opening the picture-signal channels.

The whole installation is contained in panels mounted in racks on a testing bench, a photograph of which is given in *fig. 4*. On the right of the bench is the valve under test, and in the middle of the bench, at the bottom, is the metal case containing the cathode-ray tube. The observer sitting at the bench views the screen of the tube through the opening, in front of which a camera can be placed.

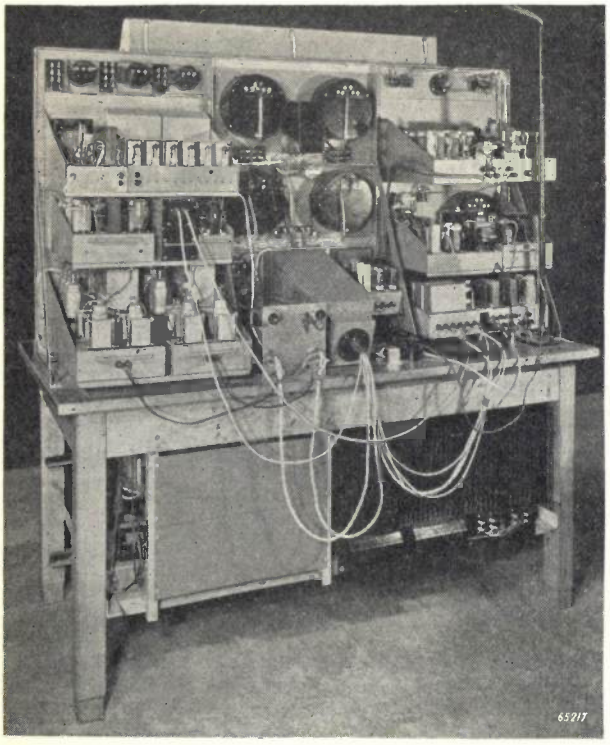


Fig. 5. Rear view of the measuring and control panels of the diagram tracer. On the right at the top the step-voltage generator for the grid voltage of the valve under test, and on the left at the top the step-voltage generator for the lattice.

Fig. 5 is a back view of the panels. Underneath the bench are the supply units and the 200 W generator for scanning the characteristic of the valve under test.

The component parts of the installation are all made as easily replaceable units, the connections between which are made with the aid of plugs and sockets. Since the whole of the apparatus contains more than 200 electronic valves, with an average life of 10,000 hours, an average of one breakdown per 50 hours has to be taken into account. Spare units are therefore provided for quick replacement.

The electronic switches

Each of the switches, as we have seen, contains three channels. In one such channel there are two valves, the switched valve T_1 and the switching valve T_2 (see *fig. 6*).

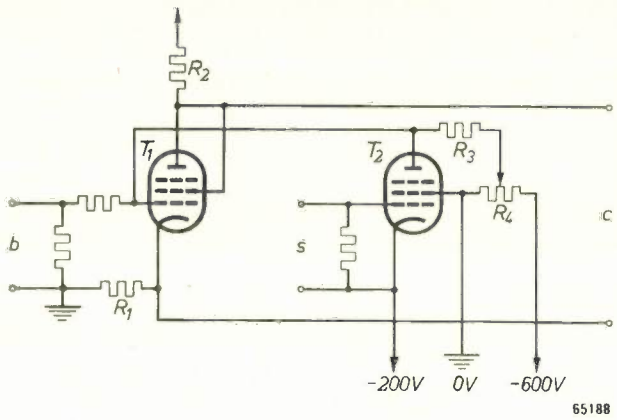


Fig. 6. Principle of the electronic switch circuit. T_1 switched valve, T_2 switching valve, b signal input, s switching-signal input, c signal output connected to the deflection amplifier.

For electronic switching a pentode is usually employed, and a positive pulse is applied to the second grid during the period that the valve has to transmit the switched signal. However, in the event of interferences arising in the switching pulses, this will affect the amplification factor of the switched valve and, therefore, the signal transmitted.

With the method employed for this apparatus the valve T_2 is cut off during the interval that the switched valve T_1 is conducting. The control grid of T_1 then receives via the resistor R_3 and the potentiometer R_4 a negative bias from a stabilized supply unit of —600 V.

During the interval that the valve T_2 is open, T_1 is cut off. The grid of T_1, which is connected to the anode of T_2, then has a voltage only slightly higher than the cathode voltage of T_2, which is

at a potential of −200 V with respect to earth. Fluctuations in the switching impulse at *s*, although strongly influencing the negative control-grid voltage of T_1, are not transmitted, because T_1 is cut off.

The correct working voltage for T_1 can be adjusted with the potentiometer R_4. The three parallel-connected valves T_1, one of which is always open and two cut off, have common cathode and anode resistors R_1 and R_2, so that the signals entering the three channels at *b* leave the switch through the common output *c*.

of negative voltage pulses lasting 1/100 second. In order to avoid hum trouble these signals are synchronized with the mains frequency (50 c/s): any hum interference arising in some way or other in the amplifier would lead to distortion of the picture on the screen of the cathode-ray tube. The signals being synchronized, any interference will only manifest itself in a slight curving of the lattice lines, which is not troublesome; if the signals were not synchronized the interferences would be manifest in a wave travelling right along the lines,

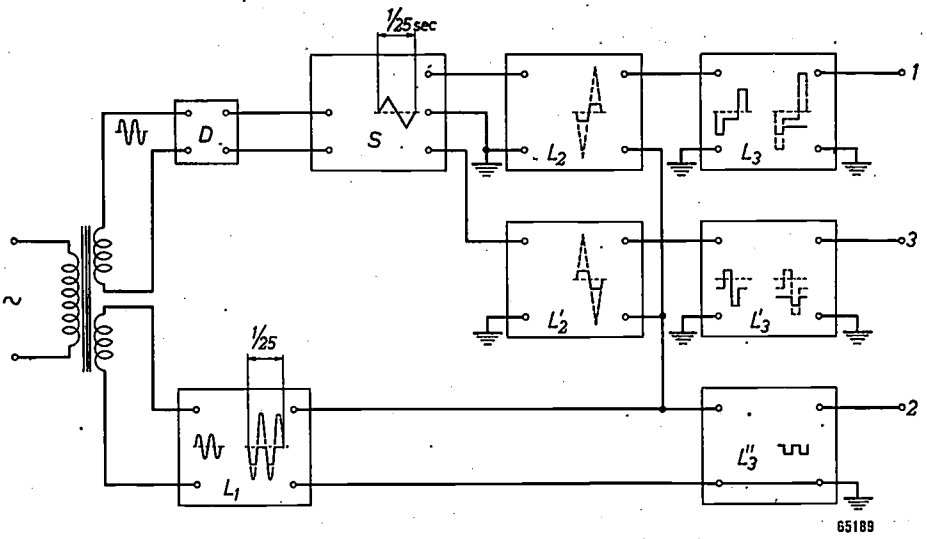


Fig. 7. Block diagram of the switching-signal generator. *L* limiters, *D* frequency divider, *S* triangular-pulse generator. Wave forms in broken lines: without limiting; fully drawn: with limiting. The output terminals *1*, *2* and *3* are connected to the switching-signal inputs s_1, s_2 and s_3 of the electronic switches.

The resistors R_1 and R_2 are of equal value, and so the output terminals receive equally large alternating voltages of opposite polarity with respect to earth. These voltages are applied via blocking capacitors to the input of the push-pull amplifier for the beam deflection of the cathode-ray tube. Apart from the fact that the two voltages in anti-phase thus obtained are suitable for feeding the push-pull amplifiers, the cathode resistance gives a heavy negative feedback highly stabilizing the gain of the valves T_1. This is necessary because in this part of the circuit the signals that are to be compared are transmitted by different valves and any variations of the gain of this stage would upset the accuracy of the calibration.

Switching-signal generator

As already mentioned when dealing with the principle of the system, the switching-signal generator supplies three different signals consisting

which would immediately be noticed, and a photograph would show thick lines unfavourably affecting the accuracy of the measurements.

The switching signals are generated in different stages, which will be explained with the aid of *fig.* 7.

A sinusoidal alternating voltage of 50 c/s is applied to a limiter L_1, and to a frequency divider *D*. The latter passes every 1/25th of a second a pulse to a circuit *S* generating a triangular voltage with a frequency of 25 c/s. This triangular voltage is applied, in anti-phase, to two limiters L_2 and L_2', which clip pieces out of the triangular voltages and thus produce square-wave pulses of the form shown in fig. 7. To these pulses are added the signals coming from the limiter L_1. When the differences between the maximum and the minimum output voltages of L_1 and L_2 (or L_2') are equal then L_3 and L_3' receive pulses of the form drawn in fig. 7. From these signals, which have a repetition

frequency of 25 c/s, the limiters L_3 and L_3' clip out negative square-wave pulses with a duration of 1/100 second.

The limiter L_3'' finally reverses the square-wave pulses supplied by L_1, so that in the second and fourth quarters negative pulses are obtained for opening the electronic switches transmitting the signals from the valve under test.

Step-voltage generator

When square-wave pulses of equal frequency but mutually shifted in phase are applied to the control grids of a number of amplifying valves with common anode resistor, then at that resistor an alternating step voltage is obtained. In *fig. 8* this is illustrated schematically for the case where there are six steps. The fundamental frequency of the square-wave pulse, synchronized with the mains, is 50 c/s, whilst each step lasts 1/100 sec. Only the ascending step voltage is used, the descending voltage being suppressed by the electronic switch.

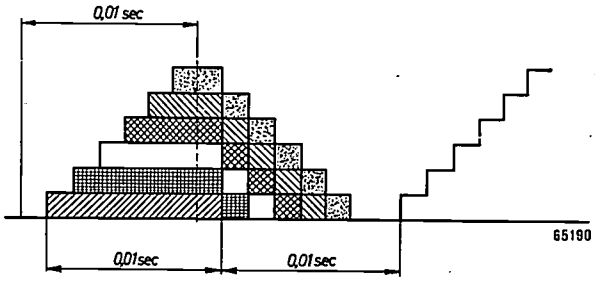


Fig. 8. The building up of a step voltage from six equal, square-wave signals, mutually shifted in phase.

The phase shifts between the successive square-wave pulses have been made equal, so that the steps of the voltages are equal in width. Thus the lines on the screen are all traced during equal intervals of time and have equal brightness.

The mutually shifted square-wave pulses are obtained, again with the aid of limiters, from sinusoidal voltages shifted in phase and generated in circuits as represented in *fig. 9* [3]). Twelve of these circuits, with a common transformer, supply twelve sinusoidal alternating voltages E_f, mutually shifted in phase. These voltages are applied to the grids of twelve valves T_1. (See *fig. 10*, where the circuits of two of these valves are drawn.) The alternating voltages E_f are much greater than the grid base of these valves (pentodes EF 40). During the greater part of the negative half of the grid voltage T_1 is cut-off; when the grid voltage is positive the valve is saturated and as a result the anode current assumes the form of a square wave. The square-wave pulses at the anode resistors R_1 are transmitted via the capacitors

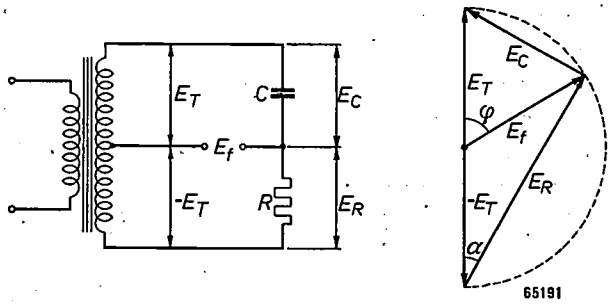


Fig. 9. Network for obtaining alternating voltages E of constant amplitude, and shifted in phase, with the corresponding vector diagram. Between the centre tap of the secondary of a mains-fed transformer and the common point of the capacitor C and the resistor R, an alternating voltage E_f is developed which, with respect to the transformer voltage E_T, is shifted over a phase angle determined by the choice of R and C. Since the vector E_C is in quadrature with E_R the extremity of the vector E_f lies on a semicircle with radius E_T, so that for all the twelve phase-shifting networks the amplitude of E_f is always equal to that of E_T. (The impedance of the transformer winding is assumed to be low.)

C_1 to the grids of twelve valves T_2 (also EF 40 pentodes) with common anode resistor R_2, amplified and then added together in R_2. The screen grid voltages of the valves T_2 can be individually adjusted between + 100 and — 10 V with respect to the cathode, for each valve separately by means of the potentiometers R_3. In this way it is possible to change the anode currents and thereby the heights of the steps, and thus the distances between the lines of the lattice. For the anode currents to be

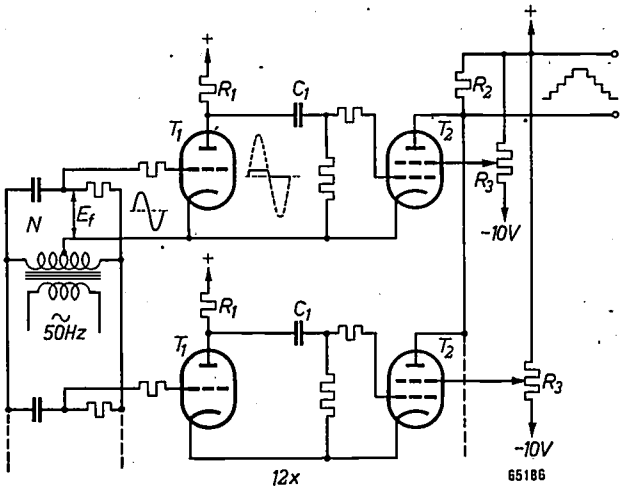


Fig. 10. Circuit of the step-voltage generator. N Network for obtaining sinusoidal voltages mutually shifted in phase; T_1 amplifying valves converting these sinusoidal signals into square-wave pulses; T_2 valves connected in parallel, with common anode resistor R_2, adding up the square-wave pulses shifted in phase, thus forming a step voltage.

[3]) The working is described, i.a., in Philips Techn. Rev. 12, 91, 1950 (No. 3).

reduced to zero it was found necessary to give the screen grid a negative potential with respect to the cathode.

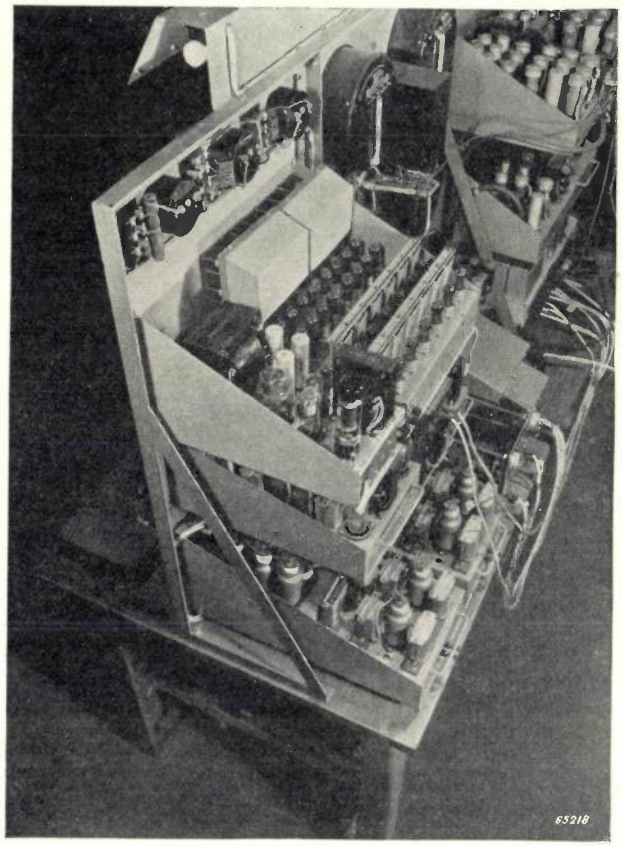


Fig. 11. Double step-voltage generator for tracing the lattice lines. Immediately behind the panel are the two sets of 12 valves of the two second stages. On the right, in two rows of six, the valves of the common first stage.

Fig. 11 is a photograph of the step-voltage generator for tracing the lattice. The circuits are duplicated so as to allow the intervals between the horizontal and the vertical lines to be adjusted independently of each other. The valves of the first stage (T_1 in fig. 10) are common to both systems, but those of the second stage are separated. For adjusting the intervals there are two sets of 12 controls, seen in fig. 4 on the right-hand panel at the top. The two controls underneath these are for adjusting the amplitudes of the two step-voltage generators, this being done without altering the proportions of the steps.

The construction of the step-voltage generator supplying the grid voltages for the valve under test (V_g in fig. 1) is similar to that of the generator supplying the signals for tracing the lattice lines. The controls for this generator are seen in fig. 4 on the left-hand panel.

Deflection amplifiers

These are designed as normal A.C. push-pull amplifiers that have to reproduce frequencies, undistorted in amplitude and phase, in a range from 25 c/s (the picture frequency) up to about 50,000 c/s (the frequency of the higher harmonics of the step-voltage generators). The valve capacitances, and stray capacitances in the output circuit, form the limiting factors, but these are compensated by means of capacitors. The usual method of reducing the influence of stray capacitances, by employing small anode resistances, could not be used here for the following reason. The output voltages required for the deflection plates of the cathode-ray tube have to be high (about 300 volts), so that if small anode resistors were used the output valves would have to deliver high powers which would be needlessly dissipated in the resistors.

By means of a compensating circuit such as used by Carpentier [4]) it has been possible to use anode resistors of a high value (100,000 ohms) and amplifying valves (type CF 50) which are capable of supplying high alternating voltages with only a small amount of distortion. The gain is constant between 10 and 15,000 c/s within $\pm$ 2% (between 10 and 60,000 c/s it is constant within $\pm$ 20%); the phase shift is less than $\pm$ 2° between 10 and 15,000 c/s.

With a push-pull amplifier the compensation can easily be obtained by coupling the grids of the amplifying valves in one of the stages, via a small capacitance C_c, to the anodes of the corresponding valves of the other phase, as is done in neutralization. The only difference is that in the case of normal neutralization only the anode-to-grid capacitance is rendered harmless, whereas here an additional voltage 90° in advance of the signal is applied to the grid via the capacitor C_c. Thus the anode current is increased by the charging current of the anode capacitance.

For a flat frequency characteristic the capacitance C_c of the compensating capacitor has to bear a certain relationship to the grid capacitance plus the stray capacitance (sum C_0), such that the condition $C_0/C_c = R_g S$ is satisfied; here R_g represents the equivalent resistance of the anode resistance of the valve in the preceding amplifying stage and the grid-leak resistance of the valve in the last stage connected in parallel, whilst S is the mutual conductance of the latter valve.

---

[4]) E. E. Carpentier, A cathode-ray oscillograph with two push-pull amplifiers, Philips Techn. Rev. **9**, 202-210, 1947.

In our case R_g = 25,000 ohms, S = 5 mA/V, C_0 = 300 pF, so that C_c has to be given a value of about 3 pF.

The compensating capacitances have been obtained with the aid of trimmers in series with a fixed capacitor of 3.3 pF.

With this system it has to be borne in mind that both the valve capacitances and stray capacitances, as well as the compensating capacitances additionally load the output stage, which at high frequencies involves an increase of the A.C. output current. It is therefore advisable to keep the stray capacitances as small as possible by careful assembly.

Supply units

For the proper functioning of the diagram tracer, direct-current sources of highly constant voltages are essential. Not only must the voltage be independent of the mains voltage and frequency, but it also has to be free of ripple and hum. Furthermore, the sources have to be variable within a wide range and have a small internal impedance.

The supply units used are in principle similar to those described in this journal in 1941 [5]). There the mains-independent voltage was obtained by balancing the output voltage against the reference voltage from a dry-cell battery; the potential difference between the two voltages influences the resistance of a regulating valve incorporated in the current circuit in such a way that the output voltage remains practically constant. In the apparatus used here the dry-cell batteries are replaced by separate voltage sources kept constant with neon voltage-reference tubes. For the functioning of the regulating valve a pentode is used instead of a triode. Since a high degree of stabilization requires a high amplification factor in the regulating valve, with a pentode a better regulation is obtained, whilst at the same time the hum ripple is suppressed to less than 1 mV. (For details see the article quoted in footnote [5]).) With this construction of the supply units a variation of 5% in the mains voltage causes a variation of only 0.01% in the output voltage.

For the supply of the deflection amplifiers, however, the hum level is still too high. To reduce this further, a very large smoothing capacitance would be required. With the aid of a so-called reactance valve [6]) this can be achieved with a comparatively small capacitor. With suitably chosen elements the valve is equivalent to a capacitance of RS times the capacitance in the grid circuit, where R is the value of a resistor connected in series with this capacitance in the grid circuit and S is the effective mutual conductance of the reactance valve. By using a pentode, a factor RS = 1800 was reached.

For the anode supply of the valve under test, in addition to a direct-voltage source V_a (fig. 1) an alternating-voltage generator G is also required for the periodical variation of the anode voltage. The generator consists of an R-C generator followed by a push-pull power amplifier with an output of 200 W, this being necessary for measuring output valves. The frequency can be varied between 700 and 1800 c/s, and is asynchronous with the mains frequency, so that the transitions from one curve to the next (at the change of the grid voltage taking place synchronously with the mains frequency) are distributed over the screen of the cathode-ray tube, leaving no visible connecting lines between them.

The R-C generator consists of a two-stage amplifier fed back by a network of resistances and capacitances. The circuit starts to oscillate at the frequency at which the input voltage of the network is in phase with the output voltage [7]).

The sine-wave voltage generated in this way is applied to an amplifier the output stage of which is formed by a push-pull amplifier with 2 × 2 valves, type EL 34, connected in parallel. At an output of 200 W the maximum amplitude is about 220 volts. The maximum variation obtainable in the anode voltage is therefore $2 \cdot 220\sqrt{2}$ = 620 V. As indicated in the block diagram of fig. 1,

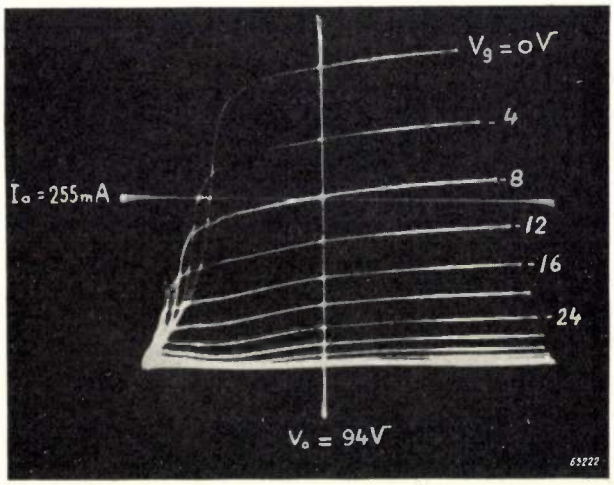


Fig. 12. Electronic crosshair on the I_a-V_a characteristics of a television pentode PL 81.

[5]) H. J. Lindenhovius and H. Rinia, A direct-current supply apparatus with stabilised voltage, Philips Techn. Rev. **6**, 54-61, 1941.

[6]) See, e.g., Philips Techn. Rev. **8**, 47, 1946.

[7]) See, e.g., J. J. Zaalberg van Zelst, Stabilised amplifiers, Philips Techn. Rev. **9**, 24-32, 1947.

in series with this generator is a direct voltage source. The two voltages can be adjusted individually, so that any anode-voltage range can be examined.

Calibration of the lattice

For calibrating the lattice the valve under test (T in fig. 1) is disconnected, and the resistor R in fig. 1, shunted across the input b_2 of the electronic switch V (fig. 2), is connected to a variable current source fitted with a milliammeter.

The voltage divider, connected to the input b_2 of the electronic switch H (fig. 2), is connected to an adjustable direct-voltage source with a voltmeter. Instead of a set of curves being traced a fluorescent spot is seen on the screen of the cathode-ray tube. With this spot, the coordinates of which are exactly known, the lattice is calibrated in the following way.

First the current and voltage sources are switched off. The spot then indicates on the screen the position of the origin of the system of coordinates. The origin of the lattice has to coincide with this point, any adjustments needed being made with the potentiometers R_4 (fig. 6) of the electronic switches.

The current and voltage of the calibrating point are then adjusted to the values corresponding to the first horizontal and vertical lattice lines to be traced. The point where these lines intersect has to coincide with the calibrating point, adjustments being made with the potentiometers R_3 (fig. 10). In this way all the lattice lines are adjusted in succession.

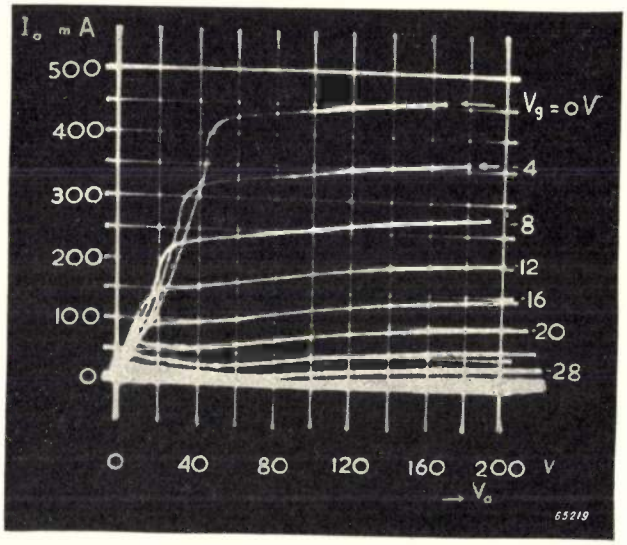


Fig. 13. I_a-V_a characteristics of a television pentode PL 81 (line-output pentode). Screen-grid voltage 180 V.

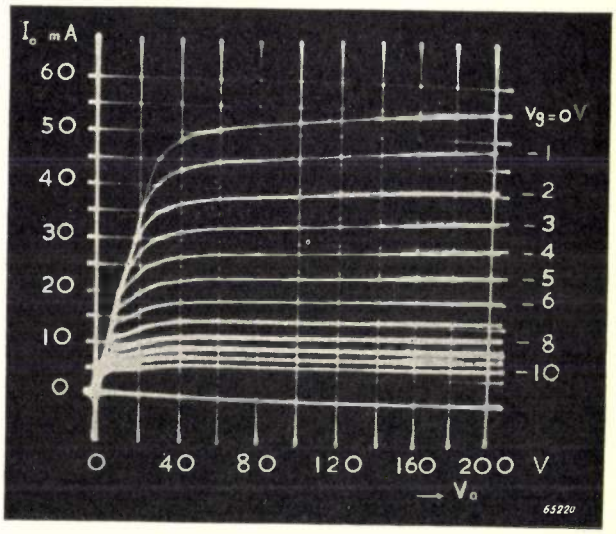


Fig. 14. I_a-V_a characteristics of the pentode part of a triode-pentode ECL 80. Screen-grid voltage 170 V.

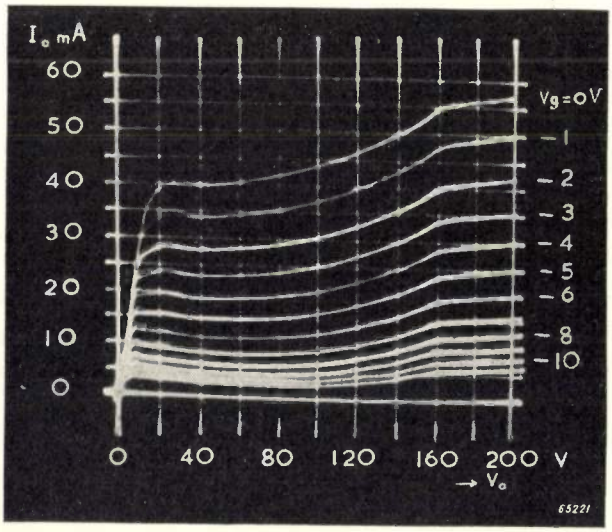


Fig. 15. As in fig. 14, but with the third grid and anode interconnected.

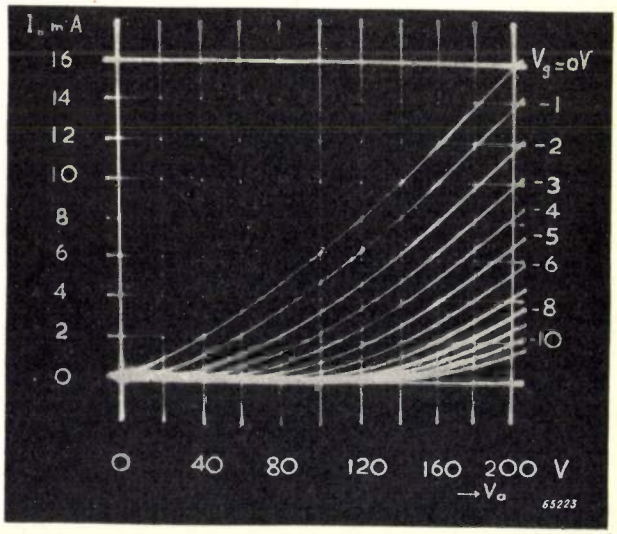


Fig. 16. I_a-V_a characteristic of the triode part of an ECL 80.

Usually the lattice will be so adjusted that the distance between two successive lines corresponds to equal potential differences at the terminal b_2.

As a rule it will be found that the voltage steps at the output of the step-voltage generator are not exactly equal, owing to the curvature of the characteristic of the operative valve of the electronic switch. Neither will the lattice lines on the screen of the cathode-ray tube then be equi-distant, because the deflection of the cathode-ray tube will not be exactly linear with the voltage on the deflection plates. However, owing to the method of calibrating described, no errors arise from these two discrepancies.

Electronic crosshair

To determine the coordinates of an interesting point, as for instance a bend or maximum in a characteristic, without interpolation between two lattice lines, the apparatus is provided with an "electronic crosshair" which can be moved across the screen. For this purpose the step-voltage generator can be replaced by variable direct-voltage sources the voltages of which can be measured with two moving-coil meters. Instead of the lattice there then appear on the screen of the cathode-ray tube two coordinate lines the values of which are known from the readings of the moving-coil meters. These meters are mounted on the panel immediately above the cathode-ray tube (fig. 4).

This method of determining the coordinates with the aid of the crosshair is less exact than the lattice method, because errors may arise from the inequality of the operative valves of the electronic switches. The crosshair method is nevertheless useful when it is desired to determine the coordinates of a point visually, for the errors due to the cause mentioned are less than those resulting from interpolation by eye between two lattice lines. Moreover, when greater accuracy is desired, there is a possibility of calibrating the coordinates of the intersecting lines, after adjustment, with the method applied for the lattice.

Fig. 12 is a photographic recording of the screen picture with electronic crosshair. *Figures 13, 14, 15* and *16* are photographs of the lattices and characteristic curves of television receiving valves, reproduced here to give an idea of the quality of the diagrams.

Summary. An apparatus is described with which a series of curves, e.g. the I_a-V_a characteristics of an electronic valve, and a calibrated lattice of coordinates can be traced simultaneously on the screen of a cathode-ray tube. In cycles of 1/25th of a sec the spot traces in succession the vertical lines of the lattice, the family of curves being examined, the horizontal lines of the lattice and once more the family of curves. The pairs of signals for tracing these three pictures are applied to the cathode-ray tube by means of two electronic switches. The lattice is calibrated by replacing the electrical quantities to be measured by two known direct voltages which are adjusted so that the spot coincides with a chosen point on the lattice. In this way the influence of non-linearities in the apparatus is eliminated. The diagrams produced can be measured with an accuracy of 1% on photographic recordings. The coordinates of any point can be exactly determined with the aid of an "electronic crosshair", which can be moved across the screen of the cathode-ray tube by the adjustment of two direct voltages.

APPLYING COATINGS BY ELECTROPHORESIS

by S. A. TROELSTRA.

621.359.3:537.363

Coating by electrophoresis, whereby very small suspended particles are deposited on an electrode with the aid of an electrostatic field, is a process that has been known for a long time. The mechanism of this process has been better understood since a deeper insight has been obtained in regard to the phenomena determining the general behaviour of suspensions. By employing organic media it has been possible to develop electrophoresis into a technically useful method, which is now being applied on a wide scale in the manufacture, for instance, of radio valves.

The investigations which have led to this and which were carried out by Verwey, Hamaker and others in the Philips Laboratory at Eindhoven, were initiated by Prof. Dr. J. H. de Boer, formerly also of this laboratory.

The principle of electrophoretic coating is easily demonstrated with the following experiment. Two electrodes are immersed in a vessel containing a rather concentrated suspension of ground soapstone (a natural magnesium silicate) in alcohol. As electrodes, graphite or some metal not readily corroding electrochemically is chosen. After applying a direct voltage of some hundreds of volts for a few seconds and then lifting the electrodes out of the suspension, it is seen that in that short space of time the anode has been covered with a layer of soapstone about 1 mm thick, whereas the cathode is still bare. When the experiment is immediately repeated with the voltage reversed in polarity it is found that the deposit on the anode has completely or partly disappeared and that the other electrode is then coated. Provided the suspension is well stirred from time to time, this process can be repeated many times. The photograph in *fig. 1* shows the coating of a graphite electrode at different stages.

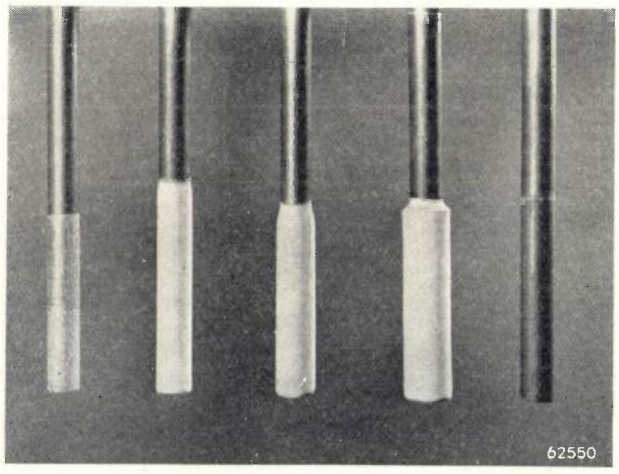


Fig. 1. Graphite electrode coated with soapstone by electrophoresis. Extreme left: after immersion only. To the right of this: after current has passed through the bath for 1, 3 and 10 seconds respectively. Extreme right: the electrode on which a deposit had been formed during 3 seconds and then removed by passing current through the bath in the opposite direction during 1 second.

Fig. 2. The settling of a suspension of quartz in water. From left to right: after 5 seconds, 10 hours and 2 weeks respectively.

This process, by means of which also insulating substances can be deposited, resembles very much that of electro-plating, where metals are deposited on the cathode electrolytically. In both cases the material to be deposited apparently migrates under the influence of the electric field, and this would indicate that the particles in the suspension are electrically charged.

The deposit formed on the electrode has very much the appearance of the sediment which after some time settles out of a suspension on the bottom of the vessel. *Fig. 2* shows what happens when, after thorough shaking, a suspension of ground quartz powder (particle diameter about 1-10 μ) in water is left to stand. The liquid gradually clears from the top downwards, but not with any sharply defined demarcation because the rate at which the particles settle differs according to their grain size. The liquid in the lower part of the vessel is

denser and the layer of sediment on the bottom gradually increases in thickness. After some time there is a firm sediment, and when this is thoroughly stirred up or the vessel is well shaken we again get the suspension in its original state.

When the quartz powder is suspended in toluene instead of in water the result is entirely different. Then, after shaking, the powder settles rapidly in the form of cohering flakes (see *fig. 3*). A clear top layer is very quickly formed, with a sharp demarcation, the interface then gradually sinking lower until at last it practically comes to rest. The precipitate shows much less cohesion than in the former

Fig. 3. The settling of a suspension of quartz in toluene. From left to right: after 5 seconds, 10 hours and 2 weeks respectively. In this unstable suspension the particles flocculate immediately and the liquid clears after a short time. Note the "buttermilk effect" on the inside of the flask.

case and with the slightest movement of the vessel tends to rise again. Furthermore, the flakes formed in the liquid cannot be entirely dispersed again even by thorough shaking, so that the original fine distribution of the powder can never be realized in this medium. In this case one speaks of an unstable suspension, as compared with a stable suspension in the former case.

Under certain conditions, about which more will be said later, also in the case of electrophoresis it may happen that the deposit shows little cohesion, so that much of the coating comes away from the electrode in thick lumps. In essence the formation of a coating by electrophoresis takes place in the same way as the sedimentation under the influence of gravity in the experiments just described. For a good understanding of what takes place in electrophoresis it is therefore of interest first to investigate the properties of suspensions in general.

Stability of suspensions

Suspensions may be stable or unstable. Let us first discuss this difference. Phenomena of stability and instability have long been known in colloidal chemistry and extensively studied with systems of particles that are microscopically indiscernable.

There is no fundamental difference between these colloidal dispersions (sols) and suspensions, and no sharp line can be drawn between the two. The main point of difference lies in the size of the particles forming the dispersed phase. For sol particles a diameter of 0.1 μ is to be taken as a maximum, whilst in a suspension about 1 μ is the minimum. Owing to this difference in size the particles in a suspension settle fairly quickly, within a few hours. In the case of sols, however, it may take days for the particles to settle, and if they are small enough they may scarcely settle at all. In fact this applies only for stable systems in which there is no mutual adhesion of the particles. In the case of an unstable system there is adhesion; aggregates (flocculates) are formed which settle much quicker than the original particles. From the rate of sedimentation alone it cannot therefore be seen whether one has separate particles or flocculates of finer particles, while account has also to be taken of the fact that the rate of sedimentation is further dependent upon the specific gravity of the settling units.

Consequently, in order to decide whether a given system is stable or unstable it is necessary to consider not only the rate of sedimentation but also other properties in which these systems differ. Here we would only mention the firmness and the volume of the sediment and the manner in which the suspension runs down a wall (the marks left on the inside of a bottle of buttermilk after it has been emptied are typical for a "flocculated" suspension).

The properties of suspensions can be explained by the assumption that there is an electric double layer around the particles. First it will be shown how this may arise and then how the stability can be accounted for by the forces due to that layer.

How the double layer is formed

Extensive investigations have confirmed that as a rule the surface of the dispersed particles is electrically charged, such in contrast to the inside of the particle, which may either be built up from neutral molecules or formed into a neutral whole by the interchange of positive and negative elements[1]).

[1]) See, e.g., J. Alexander, Colloid Chemistry, Reinhold, New York 1931; H. R. Kruyt and others, Colloid Science I, Elsevier Publishing Co., Amsterdam (at press).

In the interface it is possible for various processes to take place which cause the surface to be charged. Of the charged elements just mentioned — which also occur in a small concentration in the liquid when the solution is in equilibrium — one kind may be more strongly bound than the other. Or else ions may be formed through reaction of the surface of the original particles with the medium. And finally it is possible for foreign ions, purposely added or not, to be bound.

The surface charge, no matter how it arises, is accompanied by an equivalent but opposite charge in the liquid, in the shape of a number of mobile ions which are simultaneously subjected to the electrical attraction of their partners on the wall and to a thermal diffusion which tends to distribute them uniformly in the liquid.

The sum total of these influences is that the particle is surrounded by an "atmosphere" of counter-ions, usually referred to as a "cloud", in which the density of the charge diminishes outwards. Together with the surface charge these ions form the electric double layer previously mentioned.

A condition for the formation of the double layer is, therefore, that the liquid must have a more or less ionizing action, though it is by no means essential that the material of the particle itself should be subject to ionization.

The chemical nature of the material will usually be reflected in the structure of the double layer. Quartz (SiO_2) in water, for instance, will form on the surface silicic acid (H_2SiO_3), which dissociates according to the formula $H_2SiO_3 \rightleftarrows SiO_3^{--} + 2H^+$, whereby the two H^+ ions pass over into the medium and the SiO_3^{--} ion is left on the surface. The other possibility of dissociation: $SiO(OH)_2 \rightleftarrows SiO^{++} + 2\ OH^-$, is only found in very strongly acid surroundings (low OH concentration), so that it may be said that quartz has a decidedly negative character.

Rutile (TiO_2) is more amphoteric in nature, so that a trace of acid is already sufficient to give the surface a positive charge, such in accordance with the formula: $H_2TiO_3 \rightleftarrows TiO^{++} + 2\ OH^-$.

Zirconia (ZrO_2) has an even greater tendency to a positive charge than rutile, though here too a trace of acid in the medium is desired.

As further examples may be mentioned Al_2O_3 and Fe_2O_3. The first of these is decidedly amphoteric: with water as medium, in the surface $Al_2O_3H_2O = 2\ AlO.OH$ may be formed, the dissociation of which may take place in two ways, viz:

$$AlO^+ + OH^- \rightleftarrows AlO.OH \rightleftarrows AlO.O^- + H^+.$$

Iron oxide (Fe_2O_3) is more positive and thus dissociates according to the formula $FeO.OH \approx FeO^- + OH^-$. Under certain conditions, e.g. at a certain p_H, in suspension an amphoteric substance will have just as many positive as negative ions (or groups of ions) in the surface and thus, in the sum, it will not be charged. It is then said that the point of zero-charge or the isoelectric point is reached.

Thus it is seen that the nature of the medium will as a rule greatly influence the quantity and distribution of the ions, so that even the sign of the charge may differ. In toluene for instance there is extremely little ionization of quartz, so that there is but little development of the double layer. Finally it is pointed out that in the case of very weakly polarized substances it is often necessary to provide for a sufficient surface charge by adding highly adsorbable ions. For electrophoretic coating one mostly uses alcohols or ketones as suspension media. It is true that these liquids have a much less dissociating action than water, but on the whole the same views hold [2]).

Explanation of stability

When two suspended, charged particles approach each other the counter-ion clouds infiltrate one through the other and change in character. It is not easily seen whether the particles then attract or repel each other, but with the aid of a detailed theory Verwey and Overbeek [3]) have shown that a repulsion takes place, whereby, in the case of spherical particles for instance, the potential energy of the particles decreases roughly exponentially with their distance.

The order of the repulsion is further influenced by the size of the particles and the electrolyte content of the medium.

In addition to the repelling force determined by the structure of the double layer, however, there is also an attractive force taking part, viz. the Van der Waals-London attraction. This arises from the fluctuating distribution of the charge in atoms (also in neutral ones) and the accompanying induction of dipoles in neighbouring atoms. The corresponding attraction energy Φ_a can be represented, for not too great a distance r between two atoms (molecules), by $\Phi_a = \lambda/r^6$, where λ is a constant the value of which depends upon the properties of these atoms (molecules). When one has to do with groups of atoms or molecules the attraction energy is also influenced by the shape and size of the particles and diminishes somewhat less

[2]) E. J. W. Verwey, Properties of suspensions, especially in non-aqueous media, Rec. Trav. chim. Pays Bas **60**, 618-624, 1941; The electrical double layer of oxidic substances especially in non-aqueous media, ibidem 625-633.

[3]) E. J. W. Verwey and J. Th. G. Overbeek, Theory of the stability of lyophobic colloids, Elsevier Publishing Co., Amsterdam 1948.

rapidly with the distance from the surface of the particles.

Finally, at the very smallest distances, thus when the atoms touch each other, there is always repulsion due to the mutual penetration of the orbits of the electrons.

When both attractive and repelling forces are concerned these have to be combined to find the resultant effect [4]). From the foregoing it follows that for small distances attraction will prevail in most cases. Only exceptionally will conditions be such that repulsion predominates at any distance between two particles. The latter case is illustrated by curve *a* in *fig. 4*, where the potential energy has been plotted against the distance *r*; here we certainly have a stable suspension.

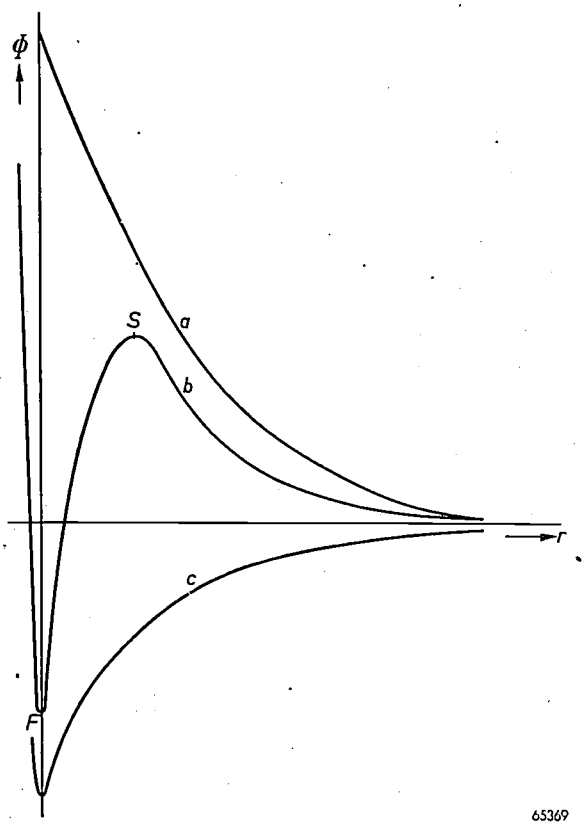


Fig. 4. Variation of the potential energy as a function of the distance between two particles,
a) for a stable suspension,
b) for a suspension which is unstable for short distances between the particles (greater concentration),
c) for an unstable suspension.

When, however, the force of repulsion is small (curve *c*) coagulation takes place spontaneously and the system is thus unstable. It is interesting to note the apparently general case that repulsion occurs at great distances and attraction at small distances; the curve *b* then shows a maximum (*S*), an "energy threshold", that has to be overcome by the partners before adhesion is possible. Thus the system then has some degree of stability according to the height of the maximum, but any incidental forces, purposely introduced or not, may cause the particles to move to the minimum (*F*) of the potential energy.

When the medium is changed, for instance by choosing a different solvent or adding or extracting some electrolyte, it is possible that the energy curve becomes transposed from the one type to the other, so that the stability likewise changes in character.

Turning back for a moment to our examples, the system of quartz in water can be described by a curve of the type *b* (fig. 4), and that of quartz in toluene by a curve of the type *c*. If ionization at the surface of the quartz is promoted by adding a little NaOH to the water, then perhaps a curve of the type *a* can be reached. It appears that in the lower alcohols and ketones there is sufficient ionization to reach an energy curve of the type *b* or even *a*.

Sedimentation in suspensions

With this rough insight into the essence of stability it is not difficult to understand why it is just a stable suspension which, under the influence of gravity, yields a dense, compact sediment: in such a stable system the particles settle separately, and those which settle most readily will temporarily stay in suspension at a distance where their weight balances with the repelling forces opposing any closer approach. (Here it is assumed that the bottom has the same sign of charge as that of the particles, and that for the interaction between particle and bottom approximately the same conditions hold as between two particles.)

The weight of the particle implies the presence of a third force, which in the energy diagram may be represented by a straight line drawn through the origin; with this as well as with the other lines the slope is a measure of the force applied [5]). As more particles settle and (via the repelling forces) bring their weight to bear upon the underlying particles this third force is increased and thus the slope of the straight line assumes a larger angle. The consequences of this are evident when we add up the original and the last energy curves. In *fig. 5* a curve is given of the type *b* corrected with a gravity term for a number of pressures directly related to the height of the sediment. It is seen that as the

[4]) H. C. Hamaker, A general theory of lyophobic colloids, Rec. Trav. chim. Pays Bas **55**, 1015-1026, 1936, and **56**, 3-25, 1937; A system of colloid phemonena, Rec. Trav. chim. Pays-Bas **56**, 727-747, 1937.

[5]) H. C. Hamaker, The influence of particle size on the physical behaviour of colloid systems, Trans. Faraday Soc. **36**, 186-192, 1940.

thickness of the layer increases so the original curve is tilted upward more and more. The distance r_M from the ordinate to the minimum M is apparently, under these conditions, the average distance of

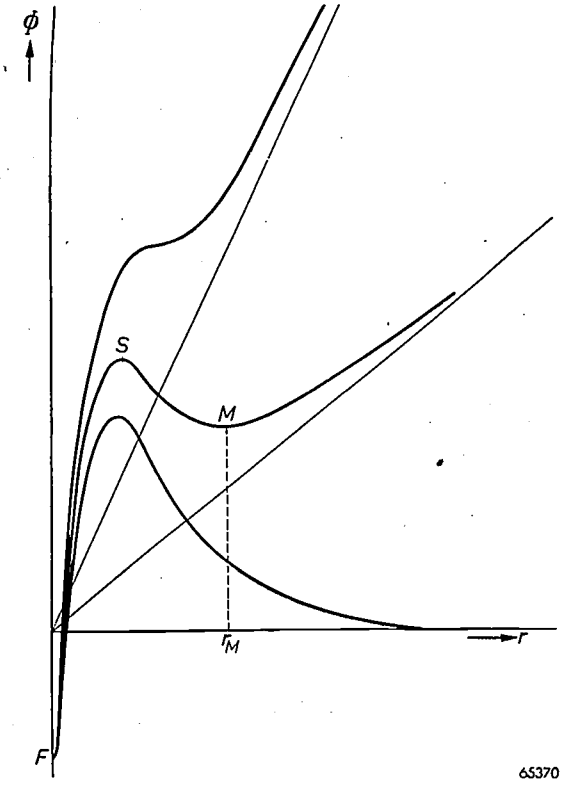


Fig. 5. As in fig. 4*b*, where account is taken of compression due to the force of gravity. The steeper the curves, the closer the minimum M and the maximum S come together and the more readily an aggregate is formed.

the particles. The difference between the maximum S and the minimum M gradually diminishes until they both disappear simultaneously. But before this situation is reached the difference S—M will have become so small that, owing to thermal motion, the particles will be propelled one after the other from M beyond S and finally reach the energy minimum F; not until then is adhesion possible. Summing up, it may be said that in a stable suspension of the type *b* first a concentrated suspension is formed near the bottom of the vessel and that this suspension gradually increases in concentration but diminishes in stability, until, starting from the bottom, the particles reach the state of coagulation.

The original curve *b* can be restored by eliminating the field of gravity — by tilting up the container — but as a rule it is not possible to separate in this way the particles that have already clung together; for this greater forces are needed, such as may be brought about by shaking or vibrating.

In the case of a sediment formed from a suspension of the type *a*, tilting of the container will indeed cause a mutual displacement of the particles: the layer is caused to flow, though often very slowly, and the particles gradually move farther apart. *Fig. 6* gives for this case the family of curves analogous to fig. 4*a*.

If, on the other hand, the sediment is formed from an unstable suspension (type *c*) then a very rare structure with little cohesion is obtained. In a stable suspension the particles had an opportunity, right up to the last moment, to unite into a very compact structure, the double layers having prevented direct contact of the particles and having made it possible for them to shear along each other; thus the double layers act as a sort of lubricant. In the case of an unstable suspension, on the other hand, partial aggregation attended by flocculation takes place already before sedimentation is completed.

Before proceeding to make a comparison between a sedimentated layer and one obtained electrophoretically we have to consider the process of electrophoresis itself.

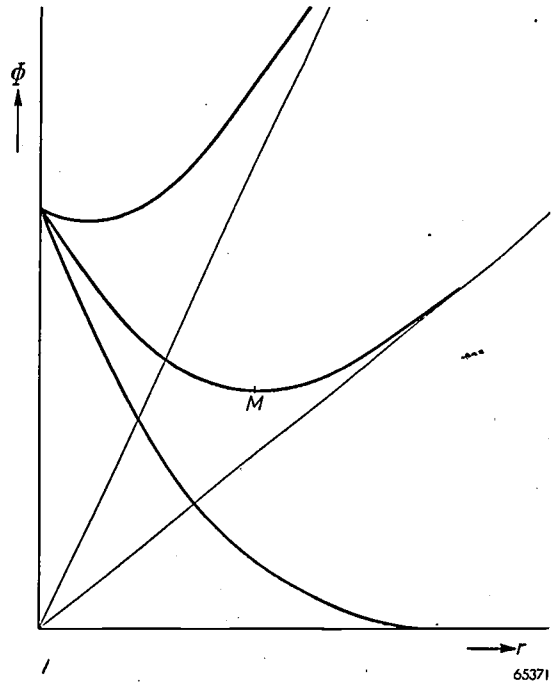


Fig. 6. As in fig. 4*a*, but taking into account the force of gravity.

Electrophoresis

Electrophoresis — the transportation of a particle in a colloidal solution or in a suspension under the influence of an electric field — has been known for a long time; in fact it yielded one of the first indications of the state of charge of colloidal particles. At first sight it seems to be a simple process: the electric force K exercised upon the charge q by the

field F can be imagined as being compensated by the friction to which the particle is subjected in the medium and which, according to a formula given by Stokes for a spherical particle, amounts to $6\pi\eta av$ (η = viscosity of the medium, v the velocity and a the radius of the particle), thus being proportional to the velocity. Hence the following formula may be written:

$$K = qF = 6\pi\eta av, \quad \text{or } v = \frac{qF}{6\pi\eta a}. \quad . \quad . \quad (1)$$

In reality the theory of electrophoresis is much more complicated than this, because, owing to the friction of the liquid carried along with them, the counter-ions moving in the opposite direction exercise a retarding action upon the particle; further, the cloud of counter-ions itself is distorted and, moreover, a thin layer of liquid is also carried along with the particle [6]).

The potential at the boundary of this liquid sheath reckoned with respect to a point in the medium at a remote distance from the particle (called the electrokinetic potential ζ), which naturally depends greatly upon the surface charge and the structure of the counter-ion cloud, has proved to be a very useful quantity for theoretical considerations, in addition to the ratio of the radius of the particle and the "thickness" of the double layer.

Calculations yield the following formula for the mobility u of a spherical particle in the suspension:

$$u = \frac{v}{F} = \frac{\varepsilon\zeta}{6\pi\eta} \cdot f(\varkappa a, \zeta),$$

where ε represents the dielectric constant of the medium and π is a quantity known from the theory of electrolytes. $1/\varkappa$ corresponds fairly well to the effective "thickness" of the double layer, i.e. the distance from the surface at which, for the sake of simplicity, any diffuse counter-ion charge can be imagined as being concentrated in one plane.

For large particles, as in our suspensions, the function $f(\varkappa a,\zeta)$ assumes the value of 1.5, so that we may write:

$$u = \frac{\varepsilon\zeta}{4\pi\eta}.$$

With the aid of simplifying assumptions it is possible to derive from this a formula of the type (1), viz:

$$u = \frac{q}{4\pi\eta\, a(1 + \varkappa a)}.$$

By way of orientation it may be said that the value of a will usually lie between 10^{-4} and 10^{-3} cm and that of $\varkappa$ between 10^4 and 10^6 cm^{-1}, so that the velocity is often much smaller than the value that would follow from equation (1).

Now, when a potential difference is applied between two electrodes immersed in a suspension, a current begins to flow through that suspension and the suspended particles migrate towards one of the electrodes (which one depends upon the sign of the charge; to define our thoughts it will be assumed that q is negative and thus that the particles migrate towards the anode).

Electrophoretic coating

Having explained in detail the formation of a sediment under the influence of a gravitational field, we can by brief in dealing with the deposition of a coating by electrophoresis. The mechanism of the process is on the same lines, except that the motive force is of an electric nature [7]). The same curves as those given in figs 5 and 6 apply for the variation of the potential energy, but curve b will be considered more closely in a quantitative sense.

In *fig.* 7 the potential energy of two small spheres has been plotted against the distance between their

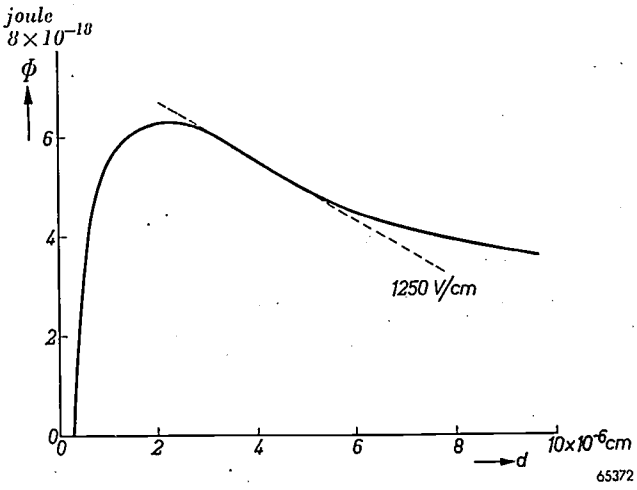


Fig. 7. Potential energy of two small spheres with surface charge and ion cloud, as a function of the distance d between their surfaces. For the numerical calculation the values taken are: $\varepsilon = 26\ \varepsilon_0$F/m [8]) (acetone), $\zeta = 100$ mV, $a = 10^{-4}$ cm, $\varkappa = 10^5$ cm^{-1}, whilst a plausible value has been given to the Van der Waals-London constant.

[6]) J. Th. G. Overbeek, Advances in Colloid Science III, 97-135, New York 1950.

[7]) J. H. de Boer, H. C. Hamaker and E. F. W. Verwey, Electro-deposition of a thin layer of powdered substances, Rec. Trav. chim. Pays Bas **58**, 662-665, 1939.
H. C. Hamaker and E. F. W. Verwey, The role of the forces between particles in electro-deposition and other phenomena, Trans. Faraday Soc. **36**, 180-185, 1940.

[8]) Here units of the rationalised Giorgi system are used. ε_0, the dielectric constant in vacuo, amounts to $10^7/4\pi c^2 = 0.885 \cdot 10^{-12}$ F/m, where c represents the velocity of light in vacuo.

surfaces. The constants determining repulsion and attraction have been given plausible values, which lead to a curve of the type of fig. 4*b*. It is observed that to the right of the maximum the slope of the energy curve indicates the magnitude of the repelling force exercised by one particle upon another, thus also the magnitude of the force which is just sufficient to prevent the particles from moving apart.

The slope in the bending point of the curve is proportional to the minimum force required to drag the particle up against the "energy crest". The field strength F needed for this proves to be 1250 V/cm. The question arises whether it is possible to produce this relatively high field strength.

In principle, of course, this is always possible by raising the voltage between the electrodes. A high field strength can also be obtained locally, however, with a relatively small potential difference, namely in the vicinity of a cylindrical electrode of small diameter.

We shall now calculate the field for an arrangement with two coaxial, cylindrical electrodes. Ignoring boundary effects, it may be assumed that the field is dependent only upon the distance r to the axis of the cylinders and that it does not change in the direction parallel to the axis. Denoting the specific conductivity of the suspension by γ and the height of the cylinder by l, the current I passing along a cylinder at the distance r from its axis is given by

$$I = 2\pi r l \gamma F . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In a homogeneous suspension the value of γ is the same everywhere, so that apparently F is inversely proportional to r.

By integration we find the potential E as function of the distance r to the axis as:

$$E = \int_{R_1}^{r} F \, dr = \frac{I}{2\pi l \gamma} \log_e \frac{r}{R_1} \quad . \quad . \quad . \quad (3)$$

where R_1 represents the radius of the inner electrode.

In *fig. 8* the potential and the field strength have been plotted for a potential difference of 200 volts between cylindrical electrodes with radii of 0.05 and 1.25 cm. With these dimensions the field strength at the surface of the inner electrode appears to be 1240 V/cm.

This is thus the potential variation when the conductivity γ has the same value everywhere in the vessel.

The electrophoresis itself does not cause any local differences in concentration of the particles in the suspension, not even in a cylindrical arrangement, as one might at first sight be inclined to suppose. Given that the concentration of the particles is c, then the number of them passing along a cylinder at a distance r from its axis is:

$$2\pi r l \cdot uFc. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Now, according to equation (2), F was inverscly proportional to r, whilst the mobility u has to be taken as being constant. Thus the number of particles passing along is independent of r and in a cylindrical shell there are just as many particles entering from the outside as there are emerging from the inside, so that there is no increase in the number of particles in the shell.

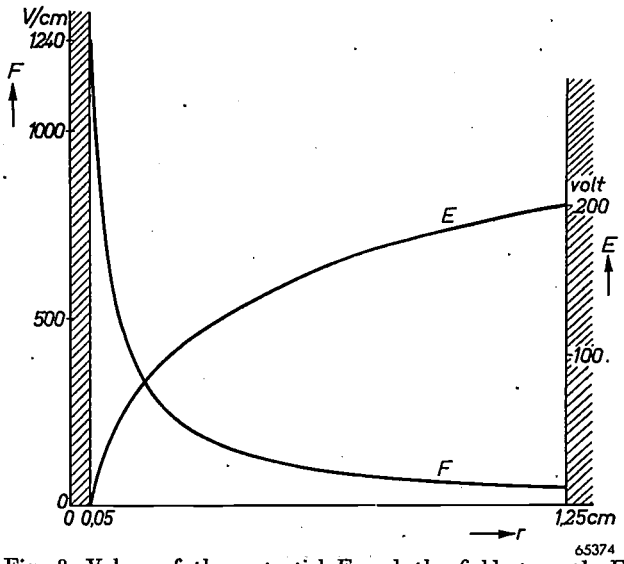


Fig. 8. Values of the potential E and the field strength F in the space between two coaxial cylinders, as functions of the distance r to the axis, at a voltage of 200 V between the cylinders.

Now in many cases it has been observed that, in spite of stirring, with a constant electrode voltage the current gradually decreases. Apparently, therefore, the deposited layer has a higher resistance than the suspension, and as a consequence the voltage drop within the deposit increases.

Curve *2* in *fig. 9* gives the potential variation inside the vessel when assuming that around the inner electrode a layer has been formed in which the conductivity is five times smaller than that in the suspension.

There are indications that the resistance of the layer is not a simple dissipative resistance. Considering that, owing to the electric field, a change takes place in the counter-ion atmosphere around each particle, which amounts to a polarisation, it is not surprising that often a much lower resistance is found with alternating current than with direct current. All the factors related to the resistance are, however, by no means explained.

As is the case with sedimentation, the forces acting upon the particles in the layer are continually increasing, since the outermost layers bear upon the underlying ones. But with electrophoresis the

forces may also, in certain cases, be considerably increased by the gradually increasing field strength inside the layer.

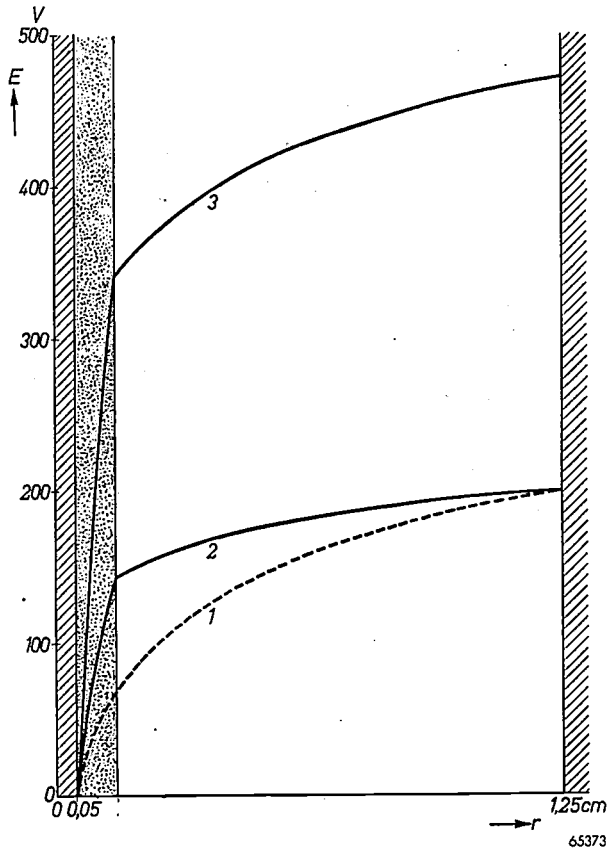


Fig. 9. Value of the potential E in a cylindrical arrangement for electrophoretic experiments. *1* when applying an electrode voltage of 200 volts; *2* when a layer 0.1 cm thick has been formed, while keeping the electrode voltage constant; *3* ditto when the current is kept constant and the voltage increases owing to greater resistance.

The electrophoretic yield

When a layer is being formed by means of electrophoresis the particles do not lose their charge upon being deposited on the electrode, just as is the case when the particles settle on the bottom of a vessel owing to the force of gravity. The process of deposition may be regarded as an accumulation of, for instance negative, particles on the electrode (anode), whereby the current transportation corresponding to their displacement (Kohlrausch's law) is brought about by secondary processes at the electrode. If the material of this electrode is not a precious metal like gold or platinum, e.g. of copper or nickel, then it yields new ions (Cu^{++} or Ni^{++}) which neutralize the supplied charge in the vicinity of the electrode. If the electrode is made of a precious metal or graphite then the ions (e.g. H^+) formed from the solvent take over this function, whilst their partners (OH^-) are discharged at the surface.

Thus it is just the same as in the case of the accumulation of sulphate ions (SO_4^{--}) in the electrolysis of a sulphate solution. On an anode of platinum a solution of sulphuric acid (H_2SO_4) is then formed. The only difference is that by the same process the particles in a suspension cake together into a layer like "giant ions", whilst the neutralizing charge — thus the counter-ion cloud — finds plenty of space between the particles. The "thickness" of the double layer may be of the order of 10^{-6} cm and the size of the particles 10^{-4} cm.

The perceptibility of the layer suggests a likeness to a precipitated layer of a metal. But in that case atoms are formed from the ions by primary discharge and the result is a solid phase with entirely different properties.

In the case of a simple electrolysis the amount of material deposited, A, in proportion to the current I is given by the constant ratio of weight and charge of the ion concerned. In the case of electrophoresis however this "equivalent weight" is a factor determined by the incidental state of charge of the colloid and also by the part played by the particles in the total charge transportation, thus by their mobility u and the sum of the degrees of mobility of all charge carriers, all multiplied by the respective concentrations. Generally, when c is the concentration of the suspended material, the amount in weight passing through the cylindrical shell per unit of time at a distance r from the axis is equal to

$$A = 2\pi l r c u F.$$

Further, according to formula (2):

$$I = 2\pi l r \gamma F,$$

so that $A/I = cu/\gamma$, where $u = \varepsilon\zeta/4\pi\eta$.

From this formula it appears, as already observed by Hamaker [9]), that the amount deposited per unit of time is proportional to:

1) the mobility u of the particles, a quantity which characterizes the suspension used;
2) the concentration c of the suspended particles;
3) the current strength I, which in turn is proportional to the surface $2\pi rl$ of the electrode.

Further the quantity deposited is inversely proportional to the conductivity γ of the liquid. It is therefore of importance to avoid as far as possible any excess of electrolyte — i.e. ions other than

[9]) H. C. Hamaker, Formation of a deposit by electrophoresis, Trans. Faraday Soc. **36**, 279-287, 1940.

those belonging to the particles and their counterions — without, however, reducing too much the charge on the surface of the particles (influence on ζ!).

When preparing suspensions of, for instance, ZrO_2, ThO_2 or Al_2O_3, the powder is often boiled with hydrochloric acid in order to increase the number of ionized groups in the surface. This has to be followed by a thorough washing out of the excess of electrolyte.

In the foregoing we have seen how, in the case of suspensions of the type represented by curve *b* in fig. 4, a compact deposit is formed. With suspensions of the type *a*, where the particles do not fall into a potential trap — thus very stable suspensions —, it may be expected that after the disappearance of the electric field the particles will move away from each other again. And indeed in such a case it is sometimes seen that the layer gradually runs off the electrode in the form of thick drops. In other cases, however, the internal friction in the layer is so great that, owing to evaporation of the solvent, a certain firmness is reached before any noticeable change takes place. Generally speaking, therefore, both suspensions of the type *a* and those of the type *b* are suitable for electrophoretic coating.

Other factors playing a part in electrophoretic coating

Suppression of the formation of gases

One might ask why preference is not always given to aqueous suspensions when a high degree of stability is required, considering that in these systems the double layer is developed better than in alcohols or acetone, media which were mentioned at the beginning of this article.

It appears, however, that in aqueous media, in the presence of the necessary field strength, electrolytic decomposition causes a very troublesome development of gas, which leads to a pitted layer and may even render deposition quite impossible. The organic liquids mentioned, while still having sufficient polarity, have the advantage that they act at the same time as a depolarizer, that is to say, by chemical reaction they are able to suppress the development of both H_2 and O_2.

But even when organic media are used, in spite of the low degree of solubility of most salts, the coating may be pitted if too much electrolyte has been added (e.g. to ensure stability).

Drying, adhesion and shrinkage

Another advantage of organic solvents is the fact that they evaporate quickly, so that within a few seconds of the coated object being removed from the bath the layer already has a certain degree of firmness.

A question of technical importance is whether the layer shrinks in the process of drying and whether such shrinkage takes place uniformly or is accompanied by cracking. Cracks arise from the action of the surface tension of the liquid in the capillary cavities of the layer.

With a view to lending more strength to the layer and also to prevent cracking, in many cases a binder is added to the suspension, e.g. a nitrated cellulose, which makes it much easier to handle the layer. At the same time this ensures good adhesion of the layer to the carrier. For the sake of simplicity this adhesion has not been dealt with as a separate problem in the foregoing considerations. Actually there is a difference compared with the mutual adhesion of the particles, but it is only by way of exception that special measures are called for.

Structure of the layer

In point of fact the firmness of the layer, and thus also shrinkage, depends upon the manner in which the particles are packed. It has already been explained at length how a stable suspension yields a closely packed deposit. Now the possibility is not precluded that in the event of the field strength being too high the particles may not have time to arrange themselves in this favourable position but are prematurely stacked one upon the other in the minimum *F*, as a result of which the structure of the layer will be of a flocculated type.

In some cases the structure of the layer can be influenced by increasing the field strength, and thus the forces acting upon the particles in the layer, while the layer is being formed. As we have seen when discussing fig. 9, curve *2*, the field strength in the layer increases when the electrode voltage is kept constant. If the field strength is to be increased still more then the voltage has to be raised while the layer is being formed.

Since the resistance of the bath increases with the thickness of the layer formed, it can be arranged for this to take place automatically by keeping the current in the bath constant instead of the electrode voltage. The potential variation in the bath will then be as represented in fig. 9, curve *3*, from which it is seen that the field strength in the layer increases still more than is the case when keeping the voltage constant. The current can be kept constant by including the bath in the anode circuit of a radio valve with a flat I_a-V_a-characteristic.

Applications

The investigations that led to this new insight into the problem as outlined in the foregoing, and whereby theory and practice have benefited each other, have made a dependable application of electrophoresis possible. Some examples will now be given.

In the preserves industry the tin cans used are enamelled on the inside by the electrophoresis process [10]). The cans are filled with a negative emulsion formed by the dispersion of the enamel with the aid of ammonia. The contents of the can are brought in contact with a central cathode, while the can itself serves as anode. After current has passed

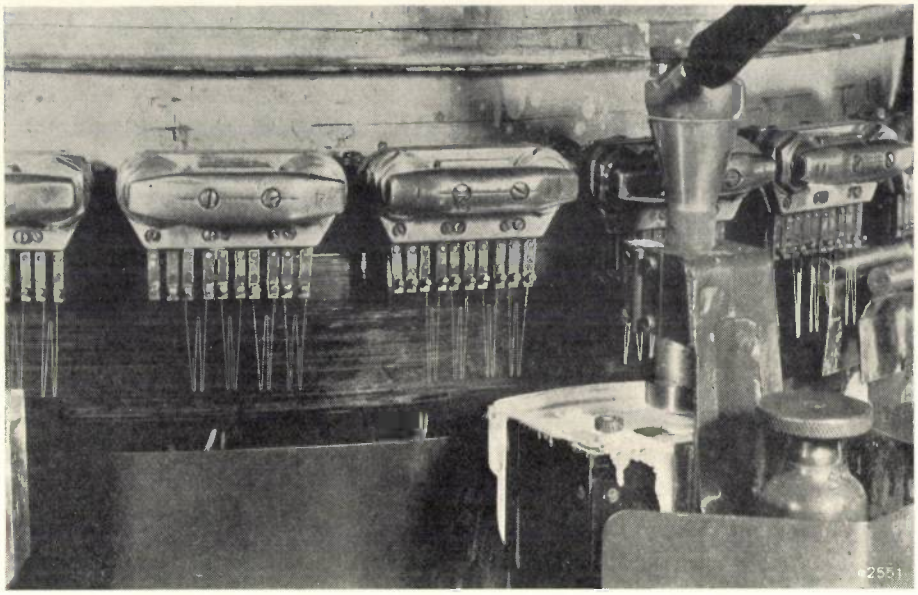


Fig. 10. The electrophoretic coating of filaments with aluminium oxide. The filaments, fixed in holders, are transported, by a turntable rotating step for step, over a tank containing the suspension. The tank is raised until the filaments are immersed, whereupon the voltage is automatically applied for the correct length of time, after which the tank is lowered. The coated filaments are then transported further into a narrow slot (on the extreme right of the photograph) where the first drying is accelerated by a current of air. In the further processing the filaments are gradually heated until the layer is sufficiently dried. This photograph was taken just at the moment that the filament holders were moved along.

In *fig. 10* a machine is illustrated as used for the electrophoretic coating of filaments with an insulating material, mostly aluminium oxide. These filaments are used for the indirect heating of oxide-coated cathodes and are shown in *fig. 11* before and after the coating.

Electrophoresis is often also employed in the manufacture of these cathodes themselves, namely for coating the metal with $BaSr(CO_3)_2$. In both cases the coating has to be further processed to make it suitable for its purpose. The Al_2O_3 is heated so as to sinter it into a firm mass. In the evacuating of the radio valve the carbonates are decomposed and thereby the oxides are formed, which at an elevated temperature emit electrons.

Another application of electrophoretic coating is found in the rubber industry, where the emulsion (latex) is deposited upon a surface and the layer is then vulcanized.

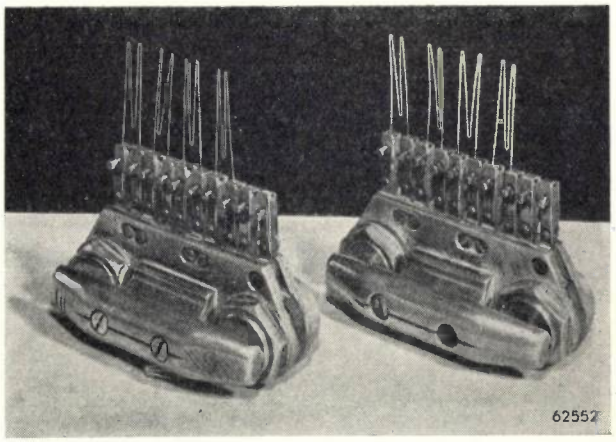


Fig. 11. Two holders with filaments, on the left before coating, on the right coated.

[10]) C. G. Sumner, The anodic deposition of oleo-resinous lacquers, Trans. Faraday Soc. **36**, 272-278, 1940.

through for a few seconds the can is emptied and rinsed out. The inside of the can is then found to be covered with a milky and strongly adhering layer of emulsion, which after drying and further treatment in a furnace has the appearance of a normal coating of lacquer. The remarkable feature about this application is that a tough liquid, difficult to handle in itself, is precipitated onto the metal from a thin, easily flowing liquid.

Summary. In the electrophoretic method of coating, microscopically small particles are deposited from a suspension onto the electrode by means of an electric field. The possibility of obtaining a good layer is closely related to the stability of the suspension used, which in turn is determined by the nature of the medium and of the substance suspended. The formation of the layer is further influenced by the shape of the electrodes, the voltage applied, the conductivity and the depolarizing power of the medium, whilst rapid evaporation of the solvent is a favourable factor.

Electrophoretic coating is applied, inter alia, on a large scale in the manufacture of cathodes for radio valves.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1944*: H. B. G. Casimir: De invloed van magnetische en electrische eigenschappen van de atoomkernen op de energieniveaux van atomen en moleculen (Ned. T. Natuurk. **16**, 198-206, 1950, No. 9). (The influence of nuclear magnetic and electric moments on the energy levels of atoms and molecules; in Dutch.)

This paper discusses the influence of the dimension of the nucleus (isotopic shift), the influence of nuclear magnetic moments and of nuclear electric quadripole moments on the spectra of atoms and molecules.

1945: W. J. Oosterkamp: The applications of Geiger counters in X-ray diffraction (Trans. Instruments and Measurements Conference, Stockholm 1949).

Report on instrumental research carried out by W. Parrish and others, Philips Laboratories Inc., Irvington on Hudson, N.Y., U.S.A., and on the author's considerations regarding accuracy in the measurement of strong lines and the detection of weak lines. Some information is given about a high-angle spectrometer and instructional applications of counter tubes.

1946: W. J. Oosterkamp: Design and applications of a dynamic electrometer (Trans. Instruments and Measurements Conference, Stockholm 1949).

Description of an instrument for measuring small direct voltages with high impedance. The voltage to be measured is converted into an alternating voltage by means of a capacitor the capacitance of which varies periodically as a function of time. This instrument serves for the measurements of α, β and γ radiation, dosimetry of X-rays and ultraviolet radiation, the testing of insulating materials, the measurement of p_H values, etc. (see Philips Techn. Rev. **10**, 338-346, 1948).

1947: N. W. H. Addink: A possible correlation between the zinc content of liver and blood and the cancer problem (Nature **166**, 693, 1950, Oct. 21).

The zinc content of various human tissues and blood has been determined spectrochemically. In the presence of malignant growth the blood has been found to contain $16\times10^{-4}\%$ of zinc (average of 6 determinations, S.D. of average $6\times10^{-4}\%$) as against $7\times10^{-4}\%$ (6 determinations, S.D. of average $1\times10^{-4}\%$) in normal cases and $12\times10^{-4}\%$ (7 determinations, S.D. of average $1\times10^{-4}\%$) in the case of non-tumorous disease, whilst in liver tissue with malignant growth $75\times10^{-4}\%$ (6 determinations, S.D. of average $12\times10^{-4}\%$) has been found, as against $53\times10^{-4}\%$ (7 determinations, S.D. of average $5\times10^{-4}\%$) in the case of non-tumorous disease.

R 141: A. van Weel: A comparison of the bandwidths of resonant transmission lines and lumped LC circuits (Philips Res. Rep. **5**, 241-249, 1950. No. 4).

In the design of resonant circuits for very-high frequencies either LC circuits with lumped elements or $\frac{1}{4}\lambda$, $\frac{1}{2}\lambda$ or $\frac{3}{4}\lambda$ transmission lines can be employed. The ratio of the bandwidths obtainable with either type of circuit has been calculated and found to depend on the product $\omega_0 Z_0 C$ (ω_0 = resonance frequency; Z_0 = characteristic impedance of the line;

C = terminal capacitance). The bandwidth of transmission lines used as circuit elements is always smaller than that of an LC circuit; the longer the line, the smaller the bandwidth. For a short-circuited $\frac{3}{4}\lambda$ line the difference is a factor 4 at least. The bandwidth of $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$ lines can be increased by introducing discontinuities in the characteristic impedance at certain points along the line. In this way the bandwidth of these lines may be made equal to that of a $\frac{1}{4}\lambda$ line. A qualitative explanation of these effects is given; this explanation holds also for cavity resonators.

R 142: W. Ch. van Geel and A. J. Dekker: Etude sur les condensateurs électrolytiques à courant alternatif (Philips Res. Rep. **5**, 250-261, 1950, No. 4). (Study on electrolytic capacitors with alternating current; in French.)

This article deals with electrolytic capacitors for alternating voltages, consisting of two plates of aluminium both covered with a layer of aluminium oxide and placed in an electrolyte. The properties of this system have been studied, special attention being paid to the potential differences between the electrolyte and the plates.

R 143: H. Dormont: Etude rhéographique des champs laplaciens à structure hélicoidale (Philips Res. Rep. **5**, 262-269, 1950, No. 4). (Study of laplacian fields of helicoidal structure by means of the electrolytic tank; in French.)

Two different methods are indicated for determining equipotentials of a helicoidally wound system of conductors by means of the electrolytic tank.

R 144: F. L. H. M. Stumpers: On a first-passage-time problem (Philips Res. Rep. **5**, 270-281, 1950, No. 4).

This report on the fluctuation problem deals with the probability that a function, starting at a time $t = 0$ from its average value E_0, will never have passed a value E_1 during the time interval $(0, t_1)$. This probability is calculated (*a*) from the Fokker-Planck equation with suitable boundary conditions and (*b*) from an integral equation derived by Schrödinger.

R 145: J. H. van Santen: Order-disorder for Coulomb forces (Philips Res. Rep. **5**, 282-287, 1950, No. 4).

According to order-disorder theories based on nearest-neighbour interaction, above the critical temperature at which long-range order is destroyed only little short-range order is left. From experimental facts (e.g. melting phenomena and order-disorder phenomena in ferrites) it is concluded that, in the case of Coulomb interaction, above the critical temperature a much larger quantity of short-range order and a corresponding larger energy of order persists. With the aid of Van der Waerden's low-temperature method it is made plausible that in the case of Coulomb forces, long-range order is much more easily destroyed for equal energy differences between the completely ordered and disordered situations than for nearest-neighbour forces.

R 146: W. Nijenhuis: Impedance synthesis distributing available loss in the reactance elements (Philips Res. Rep. **5**, 288-302, 1950, No. 4).

This paper deals with two-pole synthesis of networks in which the elements possess maximal losses. It is shown how certain difficulties arising from mutual inductances can be overcome by making use of a paper published by Bott and Duffin. In a subsequent section means are indicated how to introduce losses in coils, e.g., skin effect or eddy-currents losses. A numerical example is given in the appendix.

R 147: A. J. Dekker and W. Ch. van Geel: Les propriétés redresseuses du système Al-Al_2O_3-électrolyte soumis à une tension alternative (Philips Res. Rep. **5**, 303-314, 1950, No. 4). (Rectifying properties of the system Al-Al_2O_3-electrolyte with alternating voltage; in French.)

The Al-Al_2O_3-electrolyte system is a rectifier for alternating current. The article deals with the current-voltage characteristic of this system, mainly in the forward direction. A simple experiment shows that there is a considerable difference between the static and the dynamic response. A method is described for measuring the dynamic characteristic without any interference from the capacitive current due to the capacitance of the system. It appears that the characteristic in the forward direction shows a loop. Some properties of this loop are considered.

VOL. 12 No. 11, pp. 305-336 MAY 1951

Philips Technical Review

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

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

MOTORCAR HEADLIGHTS

by J. B. de BOER and D. VERMEULEN.

629.113.06:683.857.23

In recent years an instrument has been constructed (the "icaroscope" of Brian O'Brien) by means of which it is possible to observe the sun and the surrounding sky simultaneously, owing to the difference in brightness, the factor of which is about 10 000, being reduced to a factor of 20 with the aid of an image obtained by an intermediate link (a phosphorescent substance) with a non-linear response.

Such would be an ideal solution of the well-known problem of glare from an oncoming vehicle when travelling by night, were it not for the fact that efficiency is very low, and this applies also for other attractive solutions, such as that with polarised light. Thus, for the present, so long as there is no alternative to the systems of headlights now being used, motorists have to put up with a certain amount of glare when passing each other. In respect to this glare and its consequences from the point of view of the distances of visibility obtainable, there has been for a number of years a controversy between two schools of thought in the field of motorcar lighting, namely the European (the oldest) and the American. The object of the present article is to give the reader an insight into the most important aspects of this subject and the trend of this controversy.

On headlights in general

When walking along a road in the dark it is handy to have a pocket torch with which to illuminate the ground a short way ahead, so as to avoid stumbling over unexpected obstacles. Modern lighting of motor vehicles comes to much the same thing, only the headlights have to answer much higher demands on account of the much greater speed of travel. The "short way ahead" that is to be illuminated has to be at least equal to the braking distance, i.e. the distance within which the vehicle can be brought to a standstill. In the case of a car travelling along at a speed of 80 - 100 km/hr, as is usual nowadays on highways, under normal circumstances this braking distance is about 80 metres. Actually, however, the driver should be able to see over a much greater distance than this, because if under normal conditions he is only just able to stop the car within that distance then accidents will soon happen in the case of minor failures (brakes not properly adjusted, slippery road, speed higher than the above-mentioned limit). Moreover, the braking distance is itself increased by the appreciable time it takes the motorist to react to the perception of an obstacle. This explains why in practice by day the motorist usually keeps his eye on the road 100 to 200 metres ahead.

To be quite sure of being able to see an obstacle ahead even when the contrast between the obstacle and the road surface is very unfavourable (e.g. a darkly-dressed pedestrian on a wet road), a vertical illumination of at least a few lux, say 5 lux, is needed on the obstacle. To get this illumination at a distance of say 100 metres from two headlights the luminous intensity of each light in the axial direction has to be about 25 000 cd.

With an incandescent lamp in a parabolic reflector this is easily reached. If B is the brightness of the filament and O the area of the reflector opening, then the luminous intensity in the axial direction is

$$I_{ax} = aOB,$$

where a (< 1) is a factor allowing for the losses of light on the reflector and in the front glass. With

$B \approx 600$ cd/cm² (tungsten coil glowing at 2450 °C) and $O \approx 200$ cm² (reflector with diameter about 16 cm), before the luminous intensity drops below the said minimum the factor a may be as low as about 0.2, which would correspond to an extreme fouling or damaging of the optical system.

Of course a sufficient axial luminous intensity is not all that is required: the motorist must have a view of the road of a certain width and depth, so that the beam of light must have a certain horizontal as well as vertical spread. This leads, on the one hand, to a number of problems in regard to the construction of the bulb and the headlight, whilst on the other hand it is the source of one of the fundamental difficulties with motorcar headlights, namely the problem of the oncoming vehicle. When aiming at a sufficient spread of the beam, or, to be more exact, at the desired distribution of the light of the beam, it is almost inevitable that the luminous intensity of the headlight in the direction of the eyes of the driver of an oncoming vehicle becomes fairly great, so that both drivers suffer more or less from glare when they pass each other. Unless very drastic counter-measures are taken, such as complete separation of the traffic lanes on the road (cf. also the end of this article), this evil can only be remedied to a certain extent by sacrificing some of the desired distribution of the beam.

Now in practice, as may be presumed to be known, it is customary to use a separate beam of light when passing another vehicle. For normal travel, when there are no other sources of light on the road than those of one's own car, the "driving light" [1]) is used — a beam giving the best possible illumination of the road ahead under the limitations of power, etc. — and when approaching another vehicle travelling in the opposite direction the driver switches over to the "passing light" [1]), which is a compromise between the opposing requirements of optimum illumination and minimum glare. Of course one reckons on the other driver likewise switching over to his passing light.

Usually these two different beams of light are produced by two different filaments in one reflector. Two practical forms of this principle will be described below, where an opportunity will be afforded to deal briefly with the problems already mentioned arising in the construction of motorcar headlights. In this article, however, we shall discuss more particularly the distribution of the light in the two beams and the results that may thereby be obtained.

First we shall discuss the driving light and then the passing light.

The driving light

Statutory requirements; the 10-m screen

Let us begin with the statutory requirements that have to be met in regard to the luminous distribution of the driving light. In most European countries these requirements are practically identical, but in the U.S.A. (and also in Great Britain) they differ somewhat. They can be illustrated by imagining a screen (*fig. 1*) set up at a distance of

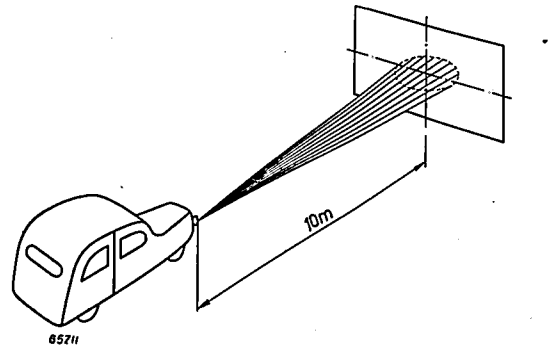


Fig. 1. To describe the property of a motorcar headlight the illumination can be shown in the points of a hypothetical screen placed at a distance of 10 m in front of a light at right angles to its optical axis. The point where this optical axis strikes the screen is the zero point of the system of coordinates drawn on the screen (see figs 2, 3, 5, 7).

10 metres in front of one headlight at right angles to its optical axis and indicating the illumination the light has to give at various points on that screen (*fig. 2*). In this diagram also a perspective image of a 6-m wide road is given, such as will be projected on the screen viewed from the headlight.

From fig. 2 it is seen that American legislators have gone much further in giving details than their European colleagues in specifying the desired distribution of the driving light, but the minimum requirements for the centre of the beam do not differ much. In the directions corresponding, according to the perspective image, to parts of the road at a distance of say 150 m, levels of illumination of 100 to 200 lux are required on the 10-m screen, corresponding to about $\frac{1}{2}$ to 1 lux per headlight at the actual distance, thus 1 to 2 lux for two lights (cf the rough calculation at the beginning of this article). The difference between the European and the American conceptions appears to lie mainly in the greater lateral spread of the beam required

[1]) Terminology proposed by the Standardization Committees of Great Britain and the United States of America for the International Standard Organization.

in the U.S.A. For practical purposes this difference is not of much consequence: also the European specifications guarantee that for normal road widths the motorist has a view of the whole width of his part of the road and also the opposite traffic lane, as well as certain strips on either side of the road (the latter is of importance with regard to the possibility of pedestrians or vehicles being about to come onto the road ahead).

go somewhat beyond the mark and — according to European feelings — too much light is directed skyward.

Beam patterns and visibility distances

However this may be, in no respect do the American and the European specifications contradict each other as far as the driving light is concerned, so that it is possible for a certain type of light to

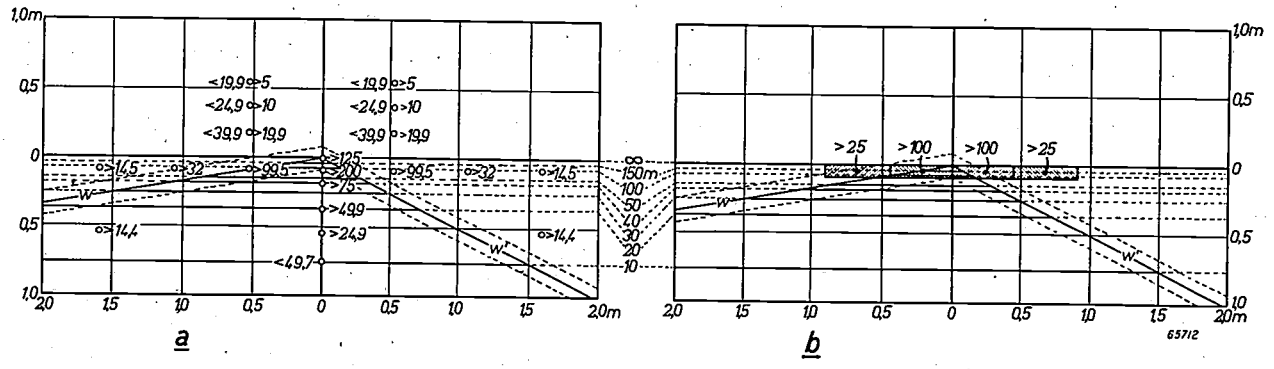


Fig. 2. Statutory requirements for driving light:
a) in the United States of America,
b) in European countries (excl. Great Britain).
The requirements are indicated by minimum or maximum values of illumination (in lux) at certain points (small circles marked respectively > or <) or within strips on the 10-m screen. The full lines *w*, *w'* are the perspective images of the left-hand and right-hand verges of the road, whilst also the projection of transverse lines on the road at different distances is indicated; the optical axis of the light is parallel to the road for these lines. The broken lines parallel to *w*, *w'* refer to the case where the optical axis of the lamp is tilted ½° upward or downward respectively.

There is a similar difference in respect to the vertical spread of the light. The requirements to be met are determined in the first place by the fact that the motorist must be able to see not only the parts of the road at a remote distance but, of course, also close by and, further, any objects which may be on or at the side of the road and which stand out above the horizon (trees, etc.). Moreover account has to be taken of the possibility of the optical axis of the headlight(s) — i.e. the axis of the beam of light — not being exactly horizontal but inclined a little upward or downward. The broken lines in figures 2*a* and 2*b* show how the road is projected on the coordinates of the 10-m screen when the angle of inclination of the lights is ½°. Bearing in mind that the chassis of the motorcar is spring-suspended and thus while the car is travelling along the road the centre lines of the headlights will swing up and down due to the movement of the chassis, it is understandable that legislators have their reasons for paying attention also to the vertical spread of the light. Perhaps too little thought has been given to this in the European specifications, whilst the American requirements answer both requirements. This is the case with the headlights commonly adopted in the U.S.A., the so called "sealed-beam" lamps (about which more will be said later). The European makes of headlights are more particularly adapted to European requirements, usually considerably exceeding the specified minima but not giving the wide lateral spread as required in the U.S.A. This is illustrated in *fig. 3*, where the beam patterns of two headlights have been given, i.e. the distribution of the illumination of the driving lights indicated by isolux lines on the above-mentioned 10-m screen. One headlight is an American sealed-beam lamp, while the other is a European make representative of the usual constructions on this continent. The much sharper focusing of the driving light from the European light is immediately noticed. To complete the comparison it may be stated that the axial luminous intensity of this European light fitted with a Philips "Duplo" bulb (for 6 V) is 60,000 cd as against 30,000 cd of the sealed-beam lamp, whilst the driving-light filament of the former consumes 35 W as against the latter's 45 W.

The extent to which a light answers its purpose

can best be denoted directly by the visibility distance, i.e. the maximum distance at which an obstacle can with certainty be perceived under certain conditions. "With certainty" signifies here: at least in a certain percentage of all cases (this percentage can still be chosen somewhat arbitrarily, say 70 or 90%). With the very special distribution of the light and owing to the influence of incidental factors such as the dancing of the beam up and down, distraction of the driver's atteniton, etc., this visibility distance is difficult to derive from fundamental data of contrast sensitivity, speed of observation and suchlike. Consequently it has to be determined directly in an experimental way with the aid of test runs under standardized conditions matched to those occurring in actual practice [2]). This applies all the more in respect to the passing lights, in the discussion of which we shall refer to the tests under driving conditions in more detail. As far as the driving light is concerned, suffice it to say that for the two types of lights of fig. 3 the visibility distances found during these test runs were 80 m (sealed-beam lamp) and 100 m (European light). These distances were the result of a large number of tests with two observers and different specimens of the two types of lights.

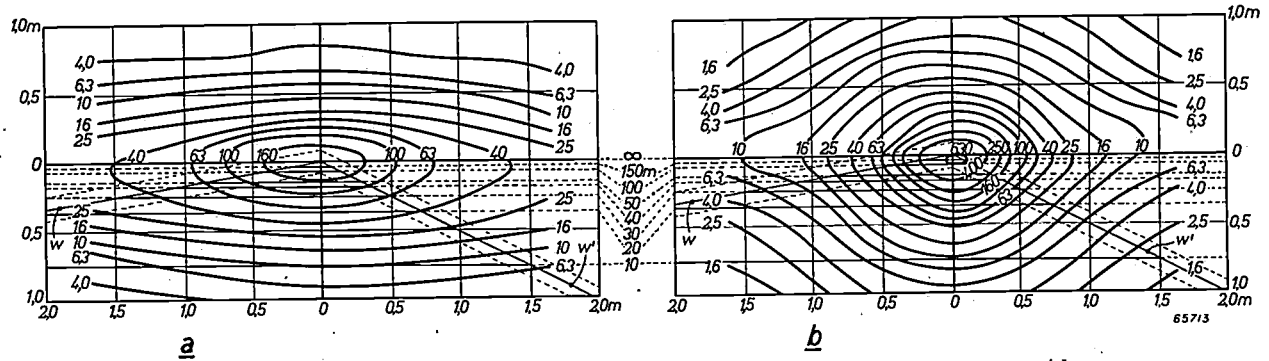


Fig. 3. Distribution of the driving light denoted by isolux lines on the 10-m screen (the lux values apply for a voltage of 6 V applied to the filament).
a) "Sealed-beam" lamp.
b) Bulb and headlight of European make.

Problems in the construction of the bulb and the headlight

How, it may be asked, are the beams of light represented in fig. 3 obtained? It would lead us too far afield to go into this question in every detail, but it is useful to touch upon the main points in the construction of the bulb and the headlight.

Typical and of importance for the construction of motorcar bulbs, of any type, is the low operating voltage from which one has to start, mostly 6 or 12 volts and in some cases 24 volts. This is related to the electric supply (dynamo with battery connected in parallel), but it is also the indicated voltage when considering the requirements that the bulbs have to meet. A certain minimum luminous intensity of the beam is desired within a certain solid angle determined by the spread required. This means that a certain luminous flux is needed, and to produce this with a not too great electrical power (no more than some tens of watts can be expended on each bulb) a type of bulb has to be used which has a high luminous efficiency (expressed in lumens per watt). A low voltage is favourable, for this implies a short and thick filament, which, for the same lifetime, can stand up to a higher operating temperature than a long and thin wire. Moreover a (coiled) short and thick wire gives a more compact filament, not only in itself (fewer turns) but more especially by reason of the fact that with a low voltage there is less danger from disruptive discharges in the gas with which the bulb is filled. Thanks to this compactness a sharply focused beam can be obtained with reflectors of moderate dimensions, so that little of the luminous flux is wasted outside the solid angle required [3]).

The short, compact coil corresponding to the low voltage is likewise favourable for making the filament strong enough to withstand the shocks

[2]) V. J. Roper and E. A. Howard, Seeing with motorcar headlamps, Ill. Eng. **33**, 417, 1938; V. J. Roper and K. D. Scott, Silhouette seeing with motorcar headlamps, Ill. Eng. **34**, 1073, 1939; H. H. Magdsick, Some factors governing visibility distance with motorcar headlights, Annex to Report No. 23b, I.C.I. 1939.

[3]) Further, the low voltage enables practically pure argon to be used for the filling of the bulb (thus without the addition of nitrogen which is necesssary for normal incandescent lamps so as to raise the disruptive voltage of the gas), so that the permissible temperature is raised and thus the luminous efficiency increased. Use is made of this in the 6-V bulbs.

and vibrations to which it is subjected when the car is travelling along bad roads or when the car is badly sprung. In passing it may be noted that this combination of shocks and vibrations is particularly serious and has constituted a difficult problem for the makers of motorcar bulbs. *Fig. 4*, illustrating a Philips motorcar bulb (of a type with only one filament), shows the shape of the filament

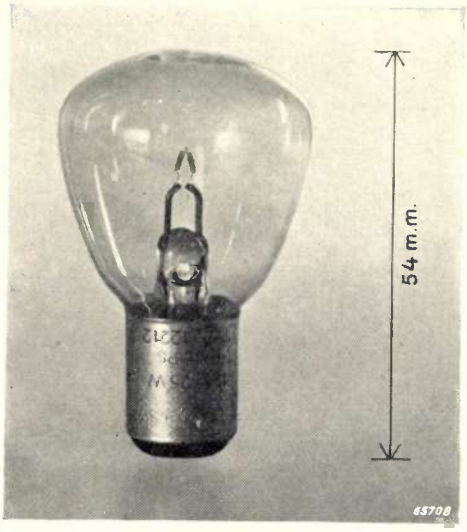


Fig. 4. Philips bulb with one filament, type 12212, for motorcar headlights.

and the supporting "mount" giving the necessary strength to the bulb. This bulb gives a total luminous flux of 640 lumens for a consumption of 35 watts.

The bulb illustrated in fig. 4 is placed in a headlight containing a parabolic reflector, in such a way that the filament coincides with the focal point. Ribs on the envelope of the bulb, on the front glass of the headlight or on the reflector itself — there are various makes, with no essential differences in the optical sense — together with the suitably chosen shape, dimensions and position of the filament ensure the right luminous distribution of the beam emitted. Owing to its sturdy construction the filament is kept accurately in position while the car is travelling along and bulbs which have been used for quite a long time still show this property.

Of course it is of great importance that when a new bulb is fitted into a headlight its filament comes to lie exactly in the right position. This is ensured by a pre-focusing of the filament in the factory with the aid of a special adjusting device [4]) by means of which the filament is given a very accurately determined position with respect to a pair of studs, stops or suchlike provided on the base of the bulb. These studs in turn determine the possition of the base in the holder, and thus within a very narrow tolerance the position of the filament in the reflector. There are numerous modifications of this principle (see, e.g., fig. 9).

Following upon this prefocusing system, it is useful to say something about the sealed-beam lamps already mentioned (see below, fig. 8). Here the intermediate link in the form of studs, etc. is dispensed with and the reflector and bulb are made into one unit, the bulb being cemented or soldered into the reflector with the aid of an adjusting device. When the bulb has burnt out both the bulb and the reflector have to be replaced by a new unit (i.e. the so-called metal-package sealed-beam). One has even gone a step farther in producing the all-glass sealed-beam lamp, in which the filament is mounted in a vacuum-sealed reflector. An advantage of both these constructions is that the problem how to exclude dust and dirt from the inside of the lamp is thereby completely solved.

The typical difference in the luminous distribution as between American and European headlights is not, however, at all correlated with the constructional differences described here. A lamp of the sealed-beam design could, in principle, be made with a European beam pattern, and vice versa.

The passing light

Statutory requirements and beam patterns

In dealing with the passing light it is again convenient to start from the statutory requirements for the luminous distribution. These are denoted on the 10-m screen in *figures 5a* and *5b*, where in addition to the perspective image of the road also the projection p is given of the line along which the eye of the driver of an oncoming car moves in a situation as shown in *fig. 6* (the dotted lines again apply for the case where the headlight is inclined $\frac{1}{2}^\circ$ upward or downward). From these diagrams it should be possible to decide what value of the illumination on the eye of the oncoming driver and therefore also on one's own eye in various situations is to be tolerated by law. Strangely enough, however, neither the American nor the European legislators say anything about this point (or rather about this line). From a comparison of figures *5a* and *5b* it can only be concluded that in directions approaching the horizontal the European lights are bound to much lower maxima than the American ones. Further it is seen that the American law stipulates that at a short distance (about 40 m)

[4]) See also: Th. J. J. A. Manders, Incandescent lamps for film projection, Philips Techn. Rev. 8, 72, 1946.

the passing light has to throw more light on the right than on the left side of the road, with a much higher minimum for the right half than is required in Europe (for the right as well as the left). For the parts of the road at distances of 50 to 100 m, which as far as observation with the passing light is concerned are much more important than distances less than 50 m, neither in America nor in Europe is any minimum illumination prescribed — again a striking omission.

is sharply cut off a little below the horizon, whilst in the American light there is asymmetry of the passing light and a comparatively gradual fall round about the horizon. The latter fact results in a much stronger glare from the oncoming car than is the case with the European light.

The asymmetry of the American passing light has the obvious disadvantage that when a car is coming round a bend to the left a great deal of light is thrown to the side of the road. What, then,

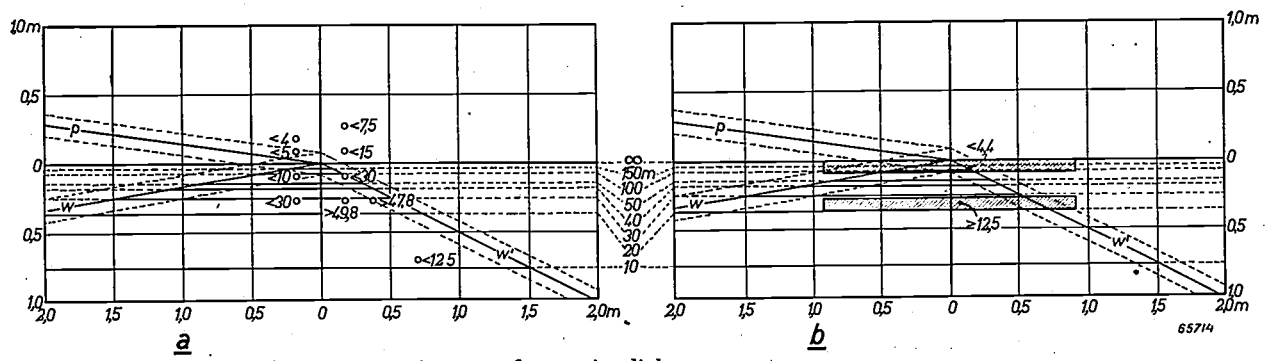


Fig. 5. Statutory requirements for passing light,
a) in the United States of America,
b) in the European countries (excl. Great Britain).
In addition to the perspective image of the road (lines w and w'), also the projectio ⁼ p is shown of the line along which the eye of an oncoming driver moves when the two cars are travelling in the middle of their own lanes on a road 6 m wide (see fig. 6). The broken lines parallel to p refer to the case where the headlight is tilted $\frac{1}{2}°$ upward or downward.

The differences between the regulations are reflected quite clearly in the actual beam patterns of the passing light used in American and in European headlights; see *fig. 7*. In the European light the beam

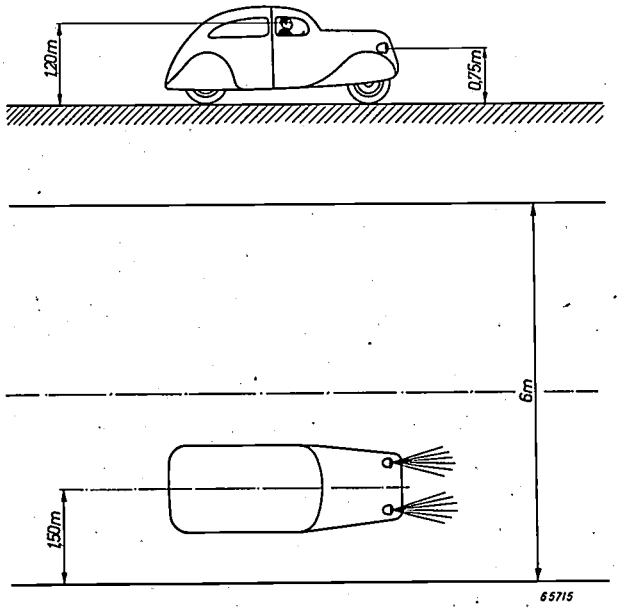


Fig. 6. Situation of an oncoming car where the driver's eye moves along a line projected from the light in line p on the 10-m screen of fig. 5.

is the reason for prescribing this asymmetry? An explanation is to be found in the remark that there is not only a bad but also a good side to the appearance of an oncoming car: two pairs of headlights then illuminate the road between the two cars. It is then obvious that each car can provide especially for the illumination of its own traffic lane, lighting this part of the road to a high level, higher than with the European system, and leave the illumination of the other lane mainly to the other car. Unfortunately in putting this idea into effect so much light is "wasted" that the advantage of the more favourable lighting of the road is at least for the greater part lost owing to the relatively stronger glare. The question whether and in how far there is any advantage left (or a disadvantage) has formed the centre of a discussion which will be dealt with presently. The difference in the two systems may be further illustrated by *table I*, showing more accurately than in fig. 7 the levels of illumination produced by the American and the European lights at various distances (a) on the road (in the middle of one's own traffic lane), and (b) on the eye of the driver of an oncoming car. For a distance of 100 m the glare and the illumination from the American passing light appear

Table I. Illumination, in lux, obtained with American and European passing lights (two headlights) at different distances between passing cars.

Distance from oncoming car in metres	American system		European system	
	on vertical object in middle of own lane beside oncoming car	on eye of driver of oncoming car	on vertical object in middle of own lane beside oncoming car	on eye of driver of oncoming car
200	—	0.04	—	0.014
150	0.18	—	0.04	—
100	0.6	0.12	0.16	0.04
50	4	0.3	3	0.1
30	—	0.6	—	0.2

away than the filament for the driving light. Furthermore this small coil is contained in a small, spoon-shaped cap. This form of mounting brings about the illustrated sharp cutt-off of the beam. At the same time the cap reflects the screened-off light, so that part of this at least can still be utilized. Special attention is drawn to the insulated arm connecting one of the poles to the end of the cap supported by an extra rod, thus giving this construction the necessary mechanical strength.

Comparison of the American and the European systems

The question whether two such widely different kinds of passing light both yield satisfactory results

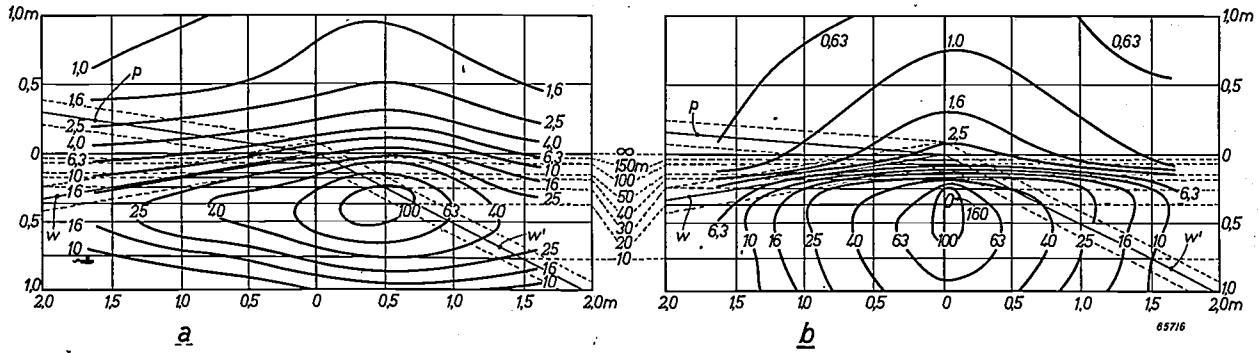


Fig. 7. Distribution of the passing light denoted by isolux lines on the 10-m screen (filament voltage 6 V).
a) Sealed-beam lamp.
b) Bulb and headlight of European type.

to be stronger by about the same factor than the levels of the European lights (assuming that the glare effect may be taken as being proportional to the illumination on the eye); for a distance of 50 m however the glare is three times and the illumination only 1.3 times as strong as that of the European passing light.

Constructional details

The manner in which the two so greatly different types of passing light are obtained can be roughly explained with the aid of *figures 8* and *9*. In the sealed-beam lamps (fig. 8) a separate filament for the passing light is placed slightly above the axis of the parabolic reflector and a little to the left (as viewed from the driver's seat) of the filament for the driving light, so that the beam of light is directed more downward and more to the right. In fig. 9 we see the Philips "Duplo" bulb type No. 6725, with which the beam pattern of fig. 7*b* was obtained. Here the filament for the passing light consists of a straight coil lying practically in the axis of the parabolic reflector but placed farther

is to be answered in the affirmative, at least in so far as countries are concerned where either the one or the other system is exclusively used. Shortly after the second world war, however, there was such a dire shortage of cars in Europe that a great many American cars had to be imported and — as has been the case with all American cars since 1939/1940 — these were equipped with sealed-beam lamps. When passing one of these American cars on the road at night the driver of a European car having the headlights common to this continent was at a disadvantage, being subjected glare from the American light, whereas the driver of the American car experienced much less trouble from glare and, moreover, found his traffic lane more brightly illuminated. This difference can be seen from table I: for the driver of the American car there were the illumination levels of the two outermost columns, while the driver of the European car had those in the two innermost columns. The ratio of the illumination on the road beside the other car compared with that on the driver's own eye at 100 metres distance was 15 for the driver of

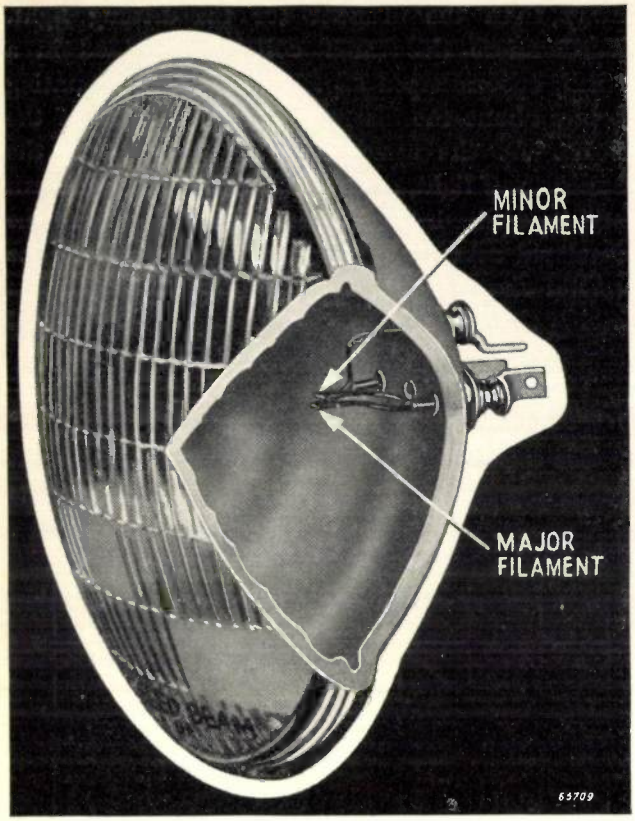


Fig. 8. All-glass sealed-beam lamp with two filaments (General Electric Co.). The filament for the passing light is above that for the driving light and displaced to the side, so that the reflector throws the light mainly onto the right-hand half of the road. The lamp bulb has not a separate, cemented-on base but three pins fused into the glass. A contact plug can be fixed on these pins. The bulb, moulded in the form of a paraboloid, is lined on the inside with a reflecting layer. (Taken from H.H. Magdsick, Ill. Eng. 35, 537, 1940.)

the American car and 1.3 for the driver of the European car!

Discussions therefore inevitably arose on the question which of the two systems was the better and thus which of the two should make place for the other in order to avoid the drivers of European cars being placed at such a disadvantage. This matter received the attention of the congress of the International Committee on Illumination held in Paris in 1948. Here we can only briefly outline the controversy, which is still not solved.

Determining the visibility distances

An impartial comparison of the two kinds of passing light will have to be based upon experimentally determined visibility distances, as previously indicated. Compared with those for determining the driving light, however, these experiments are more complicated, since when two cars are passing the visibility distance changes owing to the glare from the oncoming car gradually increasing during the approach and then, after passing, rather quickly disappearing (not abruptly, because the eye needs some time for re-adaptation).

Magdsick, Roper and others carried out such experiments (see the articles quoted in footnote [2])) already in 1938, both with the European system established since about 1930 and with the American system which in 1938 was still in course of development. As objects which had to be observed, Roper

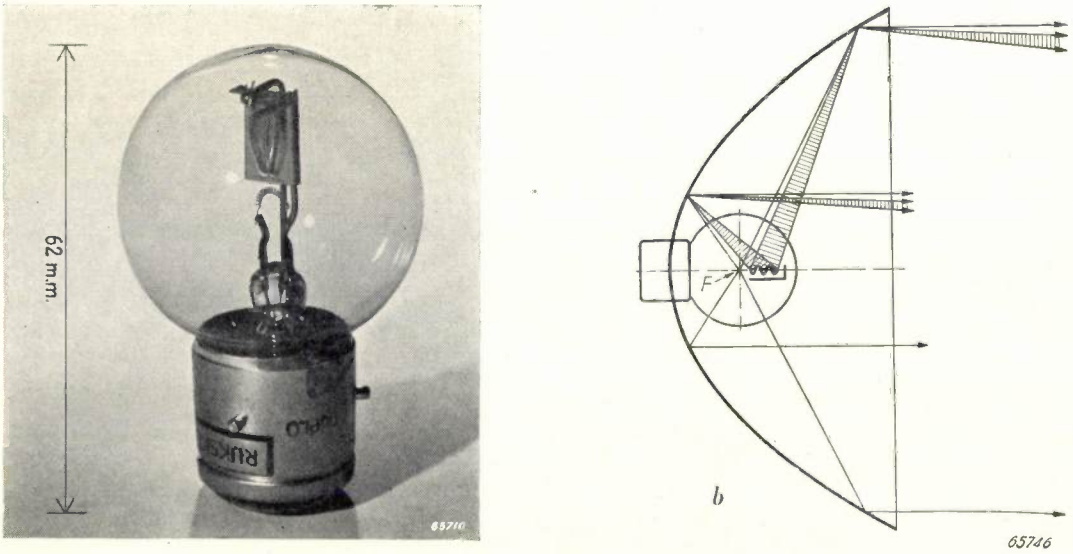


Fig. 9. *a*) Philips "Duplo" bulb, type No. 6725. On the base can be seen two of the three studs with which the lamp is focused: they ensure that the axis of the lamp coincides with that of the reflector, the filament being "prefocused" with respect to the studs. The filament for the passing light is in a spoon-shaped cap almost in the axis of the parabolic reflector in which the lamp is placed but farther away than the filament *F* for the driving light (*b*). With this arrangement the beam of the passing light is cut off fairly sharp just below the horizontal (fig. 7*b*): the cap ensures that no rays fall upon the lower half of the reflector, which would otherwise be reflected in a somewhat upward direction.

used life-size dummies set out on the right-hand half [5]) of the road (50 cm from the verge, on a road 6 m wide). Although these experiments yielded very instructive results, in our opinion the basic conditions were not very aptly chosen. With life-size dummies there is a good chance of the upper part of them being observed from a great distance owing to the silhouette effect, which Roper first remarked upon and which is to be seen in *fig. 10*. But

Fig. 10. When two cars are about to pass, a tall object on the right-hand side of the lane can sometimes be observed from a great distance owing to it standing out as a silhouette against the patch of light thrown on the road by the oncoming car. (Taken from V. J. Roper, Ill. Eng. 34, 1073, 1939.)

as the lower part of the dummy cannot be distinguished the motorist cannot judge its distance, and this is essential for deciding what action to take (swerving aside or braking). Moreover in this phase of passing a low obstacle on the road in the same position as the dummy would certainly not be noticed. Another point is that objects on the right-hand side of the traffic lane are not the most difficult to observe; with objects on the left-hand side of the traffic lane the adverse influence of glare is felt more strongly.

This criticism of Roper's basic conditions has been supported by a series of tests carried out with a stationary set-up [6]) using as objects eight moveable columns placed at equal intervals over the width of the whole road (in our case 16 m). Variations were made firstly in the distance of the oncoming car, secondly in the distance of the row of columns for each "oncoming car" distance, and thirdly in the height of the columns (22, 44, 88 or 176 cm). Measurements were taken by asking two trained observers to give the degree of visibility according to a scale of 5 steps from "invisible" to "strikingly noticeable". When all the points on the road found to have the same degree of visibility, with the oncoming headlight in a fixed position, are connected by a line we get a plan which, when plotted for different systems of headlights, permits of a comparison of their qualities. In *fig. 11* we have by way of example a series of ground plans plotted from a large number of observations for an "oncoming car" at a distance of 200 m. It is clearly seen that in this case the results with the European passing light and with object heights of 22 and 44 cm are somewhat better than those obtained with the American passing light, whereas a marked difference in favour of the American system is only found with tall objects on the extreme right of the road — exactly the conditions considered by Roper and his cooperators!

Taking as basis less tall test objects — and we believe that this is desirable — we find from the stationary tests practically equal visibility distances — also with the oncoming car at other distances

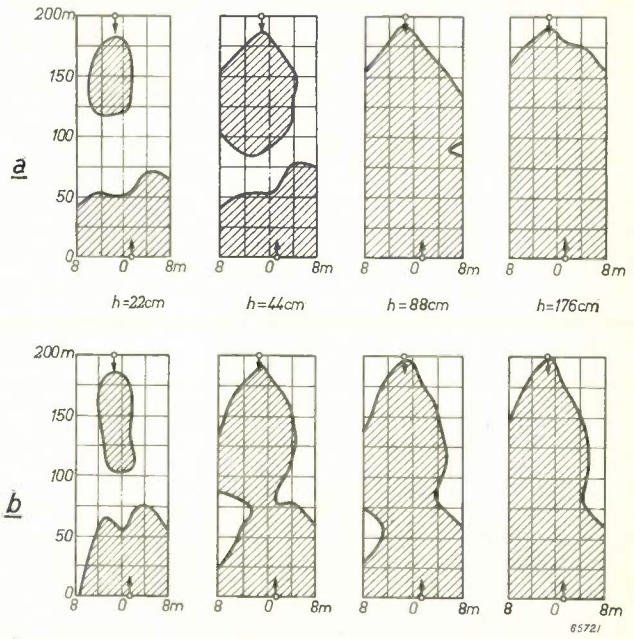


Fig. 11. Visibility distances found from tests with stationary cars placed as if they were passing each other on a road 16 m wide at a mutual distance of 200 m. The arrows denote the position and the direction of the light from the headlights. The tests were carried out with objects of four different heights h, with the passing light of American (a) and of European (b) lights with distribution of light according to figures 7a and 7b. The lines connect the points on the road where an object was given a degree of visibility "1½". The scale used for the degrees of visibility was: 0 = invisible, 1 = doubtful visibility, 2 = visible with difficulty, 3 = easily visible, 4 = strikingly visible. The hatched areas are those of at least just sufficient visibility.

[5]) In the work of Roper and others, as also in this article (fig. 2), it is assumed that cars are travelling on the right-hand side of the road, as is the rule in most countries of the world.

[6]) A report on these tests was submitted to the Paris congress of the I.C.I., viz: J. B. de Boer and D. Vermeulen, On measuring the visibility with motorcar headlighting, Congress I.C.I., Paris 1948, Appl. Sci. Res. **B2**, 1-32, 1951 (No. 1).

— for the American and for the European passing light, with a slight advantage for the latter, as illustrated by *fig. 12*.

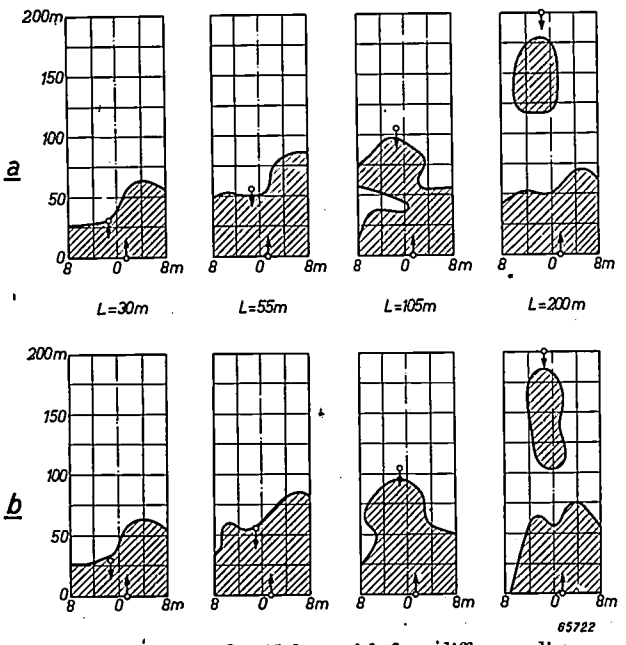


Fig. 12. The same as fig. 11 but with four different distances L between the cars; all these tests were taken with an object height $h = 22$ cm.

Although in our opinion stationary tests are quite suitable for a comparative investigation, and for this purpose they have also been applied in fact by the I.C.I., for an absolute determination of visibility distances the I.C.I. thought it advisable to carry out tests under driving conditions as well, as Roper had done but taking our objections into account. Naturally the organisation of such tests and the taking of the measurements is much more difficult than in the case of stationary tests, especially where the aim is to get results which can be internationally accepted. The I.C.I. entrusted the organisation of these driving tests to the Netherlands Foundation for Illumination, and they were carried out on the car-racing track at Zandvoort in Holland in October 1949, in accordance with directives given by the I.C.I. Committee on Automobile Lighting.

Unfortunately the Zandvoort trials did not yield any valuable results, owing to the fact that the measurements of the beam patterns of the French headlights used for the European passing light proved that they were not at all representative of the European lamps. Therefore no results of these trials will be reported here, but instead we shall give the results of driving tests carried out six months earlier by the Light-technical Laboratory of Philips at Eindhoven, which were almost entirely on the same lines as laid down by the I.C.I. for the Zandvoort trials. In particular, this implied that all conditions had to be very carefully defined and fixed. For instance, the reflection of the road surface was investigated, small irregularities in the longitudinal profile of the road were measured, the adjustment of the headlights on the two cars used was carefully checked with a special device, the beam patterns of all the combinations of headlights and bulbs used were determined, the supply voltage for the bulbs was kept highly constant and checked with the aid of calibrated voltmeters, the observers' eyes were tested in an oculistic clinic, etc.

The measuring procedure and the manner in which the results were analysed will be dealt with more fully in an appendix to this article. As to the analysis of the results, we would only point out here that there was considerable spread in the values found for the visibility distance, as is understandable. One of the most important problems in such an investigation is how to distil from the widely spread measuring results a nucleus of real differences (in so far as there are any!) [7]).

Our tests differed from the international ones only in so far that ours were limited to low test objects 22 cm high, whereas at Zandvoort tall and low test objects were used (1.50 m and 40 cm). They were carried out on two roads, a rather rough one near the Eindhoven airport and a smooth concrete road near Meteren (Netherlands). There were no systematic differences between the results on the two roads, so that they could all be used for the common purpose. The results finally worked out have been plotted graphically in *figures 13* and *14*, showing the visibility distance V as a function of the distance L between the two cars, (a) for the American passing light, (b) for the European passing light, with objects placed on the right of the road (fig. 13) and with objects placed in the centre of the road (fig. 14). Negative values of L correspond to the situation where the two cars have already passed each other, and it is then seen how the eye more or less gradually readapts itself after the glare. This visual recovery actually begins even earlier, when the cars have approached each other to within 50 - 100 m: the glaring headlights of the oncoming car have then moved a good way towards the periphery of the field of vision, where the glare effect is less. In the case where the object is on the right

[7]) The measuring procedure, the check measurements and the method of analysis, together with the results of the Zandvoort trials, are described in an extensive report submitted by the Netherlands Foundation for Illumination for the Congress of the Committee on Automobile Lighting of the I.C.I. at Turin in 1950.

of the road (fig. 13) the American passing light shows a slight superiority compared with the European, but where an object is in the middle of the road (fig. 14) the European passing light proves to be somewhat better just in the most critical phase of passing. The curves in figs 13 and 14 are based upon the measurements given by six observers with normal eyesight, and each curve is based on 70 to 80 measured points.

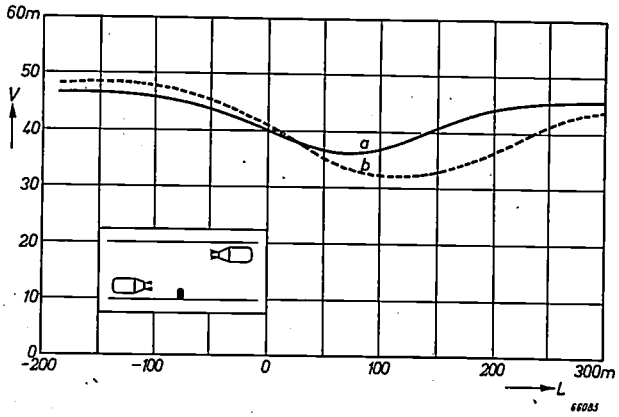


Fig. 13. Visibility distance V (defined by: observation in 50% of all cases) measured during tests with travelling cars on the Eindhoven airport and on a state highway at Meteren (Netherlands), for the passing light of American (a) and European (b) lights with luminous distributions according to figs. 7a and 7b. V has been plotted as a function of the "oncoming-car" distance L. The curves apply for an object 22 cm high placed on the right of the traffic lane.

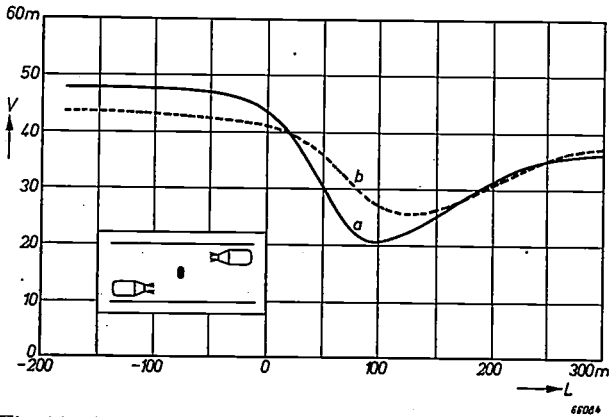


Fig. 14. The same as fig. 13 but for an object placed in the centre of the road.

Although from the results of these tests we ourselves believe that the European system is certainly not inferior to the American, the controversy between the two schools of thought in automobile lighting is still unsettled owing to the failure of the Zandvoort trials. We shall have to await further developments (the possibility of repeating the trials is being considered), but meanwhile the curves in figs 13 and 14 permit of a final general comment.

Other solutions of the problem of the oncoming vehicle

When we pay particular attention to the minimum of the curves in figures 13 and 14 we see that the visibility distances (for low objects) are not greater than 20 to 35 m. This means that one can only see past the oncoming car when this has approached within a very short distance (much shorter than the braking distance of a car). (Such an effect is also clearly seen in fig. 12.) This is the well-known phenomenon where a motorist, in passing another car, gets the feeling for a moment that he is driving into a black patch and has to trust to his lucky stars not to crash into something at that critical moment.

Motorists have grown accustomed to this hiatus in the safety measures and, so long as this "passing' situation does not occur at too great a frequency and drivers refrain from passing each other at top speed, there seems to be no cause for great anxiety. The position is different, however, on busy parts of highroads where during the peak hours there is an almost continuous line of cars passing. One is then obliged to drive practically all the way with only the passing lights on, and speed has to be reduced to 60 km/hr or less so as to adapt the braking distance to the visibility distance. To avoid this necessity of reducing speed the best solution is to illuminate such roads with sodium lamps (as has already been done for several highroads), for then the headlights can be switched off and all cars can travel safely with only the side lights on.

There are some other interesting, and more universal, possibilities of solving the problem of the oncoming vehicle. In the first place there is the use of linear or circular polarised light, already tried out in practice. When the light from the headlights of all cars is circular-polarised in a clockwise direction for instance, by means of a polariser and a $\frac{1}{4}\lambda$ foil placed in front, and the same foils are placed on the wind screen of the car, the driver can see the light from his own headlights even though it is partly depolarised in the reflection on the road. The light from the headlights of the oncoming car, however, is "seen" to be circular-polarised in an anti-clockwise direction, so that it is not really seen at all (or almost not, since owing to the dispersion the extinction is not complete for all wavelengths). There is then no longer any need to switch over to the "passing" light and one can drive on with the driving light [8]). The fact that this excellent

[8]) Among the extensive literature on this special subject there are to be mentioned a review by W. Christoph and H. E. J. Neugebauer, Z. techn. Phys. **20**, 257-264, 1939, and three reports of a more recent date: E. H. Land, J. H. Hunt, V. J. Roper, Bull. No. 11 Highway Research Board, June 1948.

solution has not yet been adopted is due mainly to the inevitable loss of light that the polarisation entails. Theoretically at most one-third but practically still less of the luminous flux from the lights is utilized, so that the headlights would consume much more electric power than they do at present.

This objection could be overcome by arranging for the driving light to be polarised not permanently but only while passing other vehicles, for instance by means of louvre-like foils being tipped downward in front of the lights and the wind screen. At the same time it should then be possible to switch over to another higher-powered filament for the driving light, thus compensating the loss of light in the foils by a momentary overloading of the battery.

Another suggestion for solving the glare problem is to dispense with the use of permanently burning headlights and to install in their place flashlamps giving periodically — say with a frequency of 50 c/s — extremely short and intensive beams of light lasting, for example, 1/10,000 sec. The eye then sees the road just as well as if it were continuously illuminated by an incandescent lamp of the same total luminous flux (apart from possible stroboscopic effects, which may make a higher frequency necessary), but it is then possible to screen off the driver's eyes, for instance by means of a rotating shutter, in the periods during which the headlights from his own car are not emitting any light. Since these periods, in the given example, cover 99½% of the whole time, the intensity with which permanently burning "foreign" light sources (e.g. of an oncoming car) are observed would be reduced to 1/200th [9]). This system has not yet been tried out in practice.

Appendix: Measuring procedure and method of analysis adopted for the driving tests

Fig. 15 is a diagrammatic representation of the road along which the trials were carried out. Two cars started simultaneously from the fixed starting points A and A'. The speed had to be uniformly increased so that the cars reached the points B and B' almost at the same moment, and from that point onward they had to travel at a constant speed of 64 km/hr. Upon reaching the points C and C' the drivers had to switch over from driving light to passing light. An object was placed on the road at a variable point between C and C'. The moment that a driver saw the obstacle he set a distance meter into action and stopped it as soon as he reached the obstacle (an assistant pulled the obstacle off the road just before the car reached it). In this way the distance from observer to object — thus the visibility distance V — was measured.

It was required to find the visibility distance as a function of the distance L between the two cars. Typical of the tests described (and also of other tests, such as Roper's) is the fact that the "independent variable" L is uncontrollable.

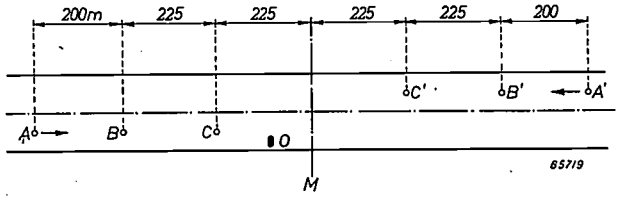


Fig. 15. Simplified driving diagram of the tests with travelling cars. The two cars started simultaneously at A and A' with driving lights on, accelerating speed up to 64 km/hr, which speed should be reached at B and B', and switching over to passing light at C and C'. O is an object that had to be perceived.

The value of L at the moment the object was observed can be deduced only from the measured value V of the visibility distance after the test has been carried out, and this is done in the following way. One starts from the assumption that from the moment that both cars have reached a constant speed they travel exactly symmetrically with respect to the point where they meet. This point is determined by one or more assistants standing along the road. If it lies at the distance P from the middle of the whole length of road travelled and the object is placed at a distance D ("object eccentricity") from that middle point, then (see *fig. 16*) apparently:

$$L = 2\,(V + D - P) \quad \ldots\ldots\ldots \quad (1)$$

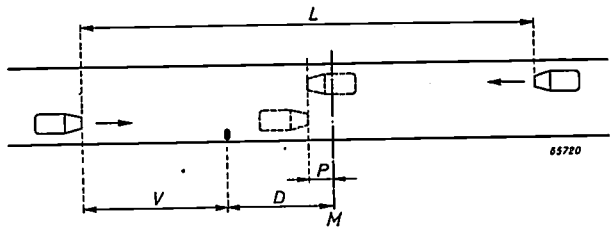


Fig. 16. The objects were placed at different distances D from the middle M of the length of the road travelled. The two cars do not as a rule meet precisely in the middle M but at a short distance P from it. From D, P and the visibility distance V it is possible to calculate the distance L between the cars at the instant the object is observed.

Suppose that for the same value of D—P a number of different visibility distances V have been recorded by different observers. These values will then have to be averaged or, to be in agreement with the definition previously given, one has to find what value is exceeded in 70% or 90% of the observations, and from the value thus found, V_0, the corresponding L_0 has to be calculated with the aid of equation (1). In this way one point of the $V(L)$ curve is determined. It is true that in practice this is not quite so simple, since owing to discrepancies in the value of P there is hardly ever a series of, nor even two, measuring points with exactly the same value of the "independent variable" D—P. But it can nevertheless be done indirectly: the measured values of V are plotted as a function of D—P and this diagram is covered with a vertical

[9]) A system of phase-control of the flashes might be desirable to avoid trouble being experienced from a passing car having the same system with, by chance, exactly the same frequency. — The idea of very short flashes synchronized with a shutter may likewise be of importance for signalling systems where it is desired to eliminate the effect of disturbing background lighting (e.g. daylight) when receiving light signals (e.g. from ground lights of an airport).

strip of a certain width. The abscisse value $D - P$ of the middle of the strip is then given, for instance, an ordinate value V_1 corresponding to that measuring point which is the next lowest above 30% of all the measuring points covered by the strip. When this has been carried out for all possible positions of the strip, the curve $V_1 = f(D - P)$ can be drawn. This indicates the distances for a 70% chance of perception and serves as a criterion for judging the passing light in question. In the same way, of course, a line can be constructed say for 90% or 50% chance of perception.

Remarkably enough, for the driving tests previously carried out neither this nor any analogous method of analysis has been applied, but a method which subsequently proved to be erroneous: for each measured value of V one proceeded at once to calculate the passing-car distance L according to eq. (1)

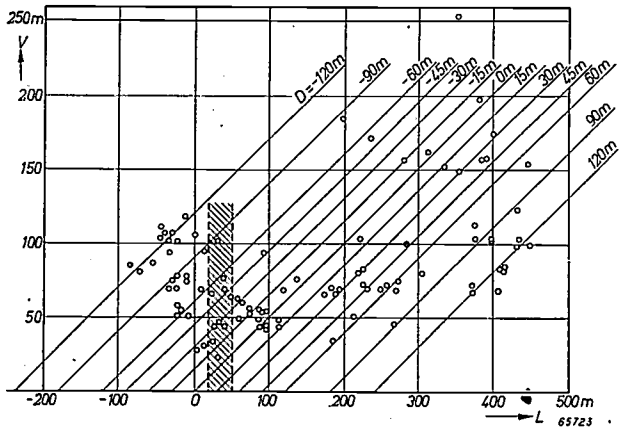


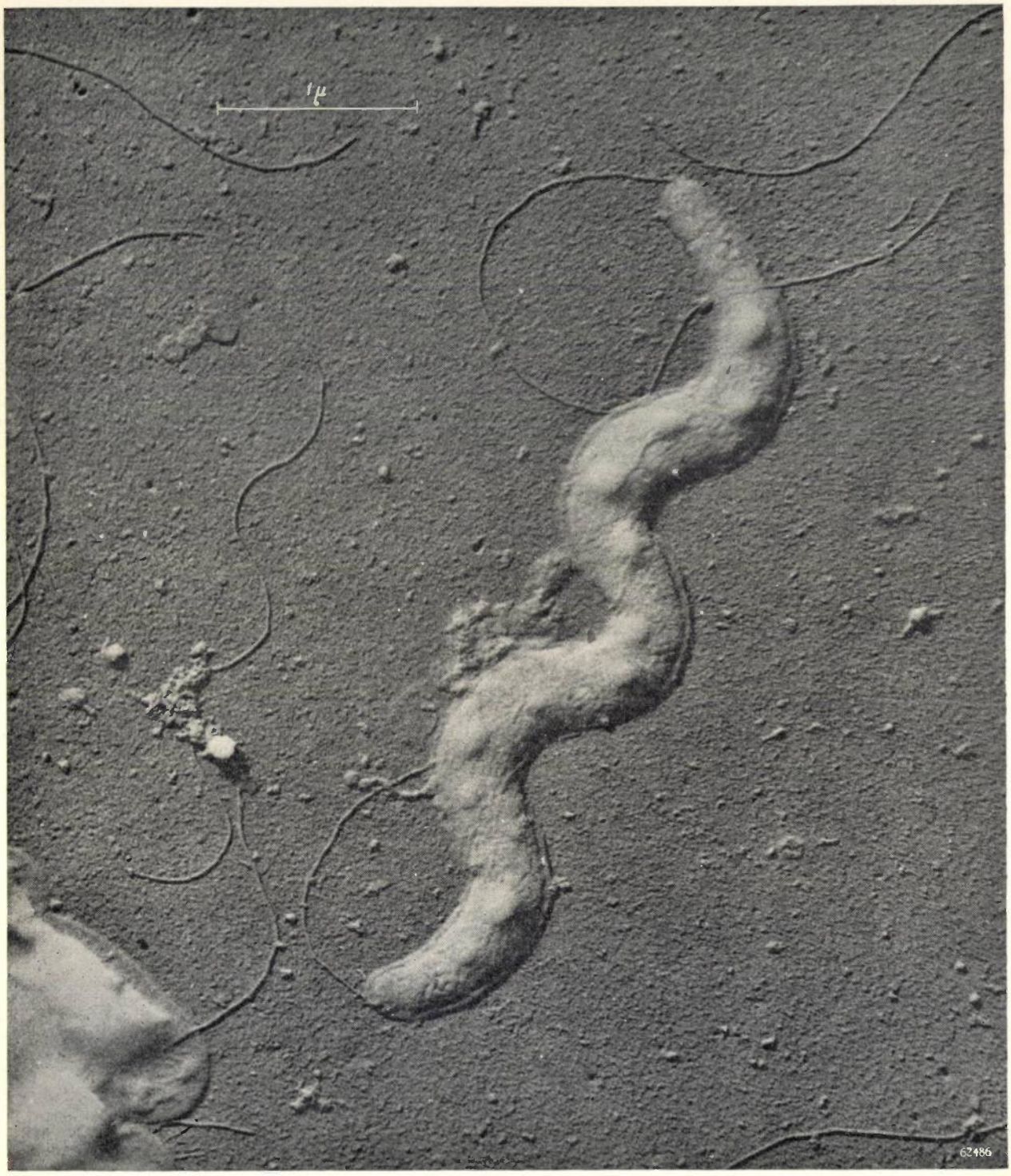
Fig. 17. Diagram explaining the method by which the "averaged" visibility distances V as function of L can be derived from the widely spread measuring results.

and then plotted all the V values as functions of L (*fig. 17*); then the method of vertical strips (or something similar) was applied to the diagram. The fault of this method lies in the fact that for a given value of L an average is taken from V values obtained with different object positions D, and thus under different conditions. The consequences of this are evident from the following. If we assume, for the sake of simplicity, that for all runs $P = 0$, then for each run $V = \frac{1}{2}L - D$, and all measuring points for runs with a certain chosen value of D would come to lie along one of the oblique lines in fig. 17, statistically distributed along it (Gauss' law of distribution). In the area of large "oncoming-car" distances, $L > 200$ m, the method of working with vertical strips will now undoubtedly yield too high values for V owing to the law of distribution in the vertical strip being unilaterally influenced: no low V values can be found in the vertical strip because there have been no tests with a sufficiently great object eccentricity D! Similarly, in the area on the extreme left of the graph one certainly finds too low V values.

It is therefore essential first to work out the relation between V and $D - P$ and then to plot this in a (V, L) graph. It is in this way that figures 13 and 14 have been constructed.

Summary: The driving light, used by a motorist when normally travelling along the road in the dark, and the passing light, to which he switches over when approaching a vehicle coming from the other direction, both have to satisfy statutory requirements in order to ensure certain minimum visibility distances and not to cause excessive glare when passing the other vehicle. The American specifications differ from those in Europe. In this article both the requirements and the actual distribution of light are graphically illustrated for the driving light and for the passing light according to the American and the European systems. The two systems are compared and some typical details of the construction of European bulbs (Philips "Duplo" bulb) and American "sealed beam" lamps are discussed. Although in several respects the statutory requirements are to be regarded as incomplete, both the European and the American systems yield satisfactory results, except at the moment immediately before passing another vehicle, for which situation there is a hiatus in the safety measures. But also apart from this, difficulties are apt to arise in the case where American and European headlights are used on the same road. In that case the motorist with European headlights on his car is greatly at a disadvantage as far as visibility is concerned. The discussions arising therefrom at the congress of the I.C.I. held in Paris in 1948 led to a repetition of the tests for directly determining visibility distances. Such tests have been carried out under stationary and under driving conditions by Philips Light-Technical Laboratory at Eindhoven, and under the auspices of the Netherlands Foundation for Illumination stationary and driving tests were likewise carried out especially for the I.C.I. at Zandvoort (Netherlands) in October 1949. Although the controversy between the American and the European schools of thought has not yet been definitely solved by these tests, the results — both of the stationary and of the driving tests — indicate that neither the one nor the other system can be regarded as being markedly superior. An interesting question related to these trials is how to analyze the widely spread results obtained from the measurements. An improved method for this is described in an appendix.

A BACTERIUM MAGNIFIED 33000 TIMES



Photograph taken with the Philips electron-microscope for 100 kV. According to its shape this bacterium belongs to the Spirillum group. It is found in water and is not morbific. The specimen has been shadowed by precipitation of an alloy of gold and palladium, thereby accentuating the relief structure of the somewhat dried-out cell. A flagellum is seen at both ends.

THE ELECTRO-ANALOGUE, AN APPARATUS FOR STUDYING REGULATING SYSTEMS

II. THE ELECTRICAL EXECUTION

by J. M. L. JANSSEN and L. ENSING.

621-52:621.3.012.8:53.072.13

The electro-analogue is a type of calculating machine specially designed for studying regulating systems. Its action is for the greater part electronic, just like that of the calculating machines destined for more universal use, a known example of which is the "ENIAC". Whilst the latter has no less than 18,000 electronic valves, in the electro-analogue described below 350 valves are used.

The name of electro-analogue is used for a collection of parts with which electrical models of automatic controllers can be built and their action studied. The component parts and the use of such an electro-analogue have been discussed in broad lines in a previous article [1]). In what follows the electric circuits will be dealt with in more detail.

In *fig. 1* the (incomplete) block diagram of the electro-analogue is reproduced for ready reference. The principal parts are:

1) the process analogue P', a network whose step-function response can be made identical with that of a process that is to be studied;
2) the model of a continuously acting controller, comprising an amplifier A, a proportional, an integrating and a differentiating element, and also the variable attenuators Z_0, Z_a, Z_b and Z_c (in this model there is also a second integrating and a second differentiating element, not drawn in fig. 1);
3) some models of discontinuously acting controllers (not drawn);
4) a collecting stage S_1, in which the output voltage(s) of the controller and a periodically repeated voltage step, which it is possible to introduce as a disturbance, are added together;
5) a collecting stage S_2, by means of which a disturbing voltage can be added to the output voltage of the process analogue;
6) a square-wave generator G for generating a square-wave voltage serving as input signal for adjusting the process analogue, or as a disturbance;
7) a cathode-ray oscilloscope *Osc* for displaying the voltages at various points of the circuit;
8) an R-C generator (not drawn) supplying a sinusoidal voltage with variable frequency for recording frequency-response curves and Nyquist diagrams;
9) a performance meter (not drawn) for measuring the accuracy of the control.

There are to be added:

10) the supply units supplying the necessary voltages for feeding the numerous valves in the above-mentioned parts.

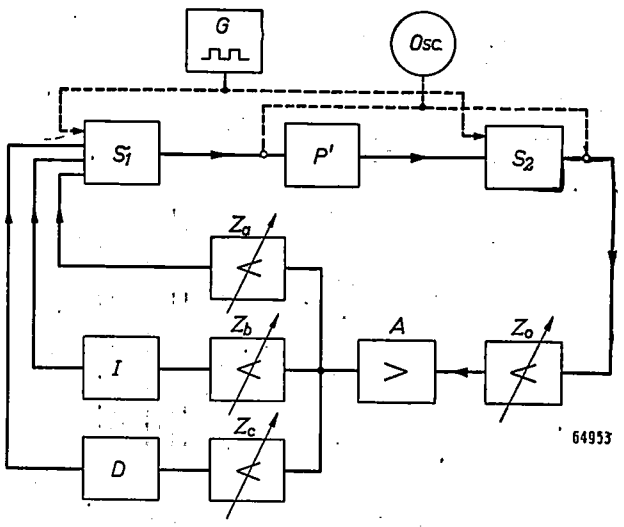


Fig. 1. Simplified block diagram of the electro-analogue. P' process analogue, with input collecting stage S_1 and output collecting stage S_2. Z_0 common attenuator, A common amplifier, I integrator, D differentiator, Z_a, Z_b and Z_c attenuators for separately adjusting the proportional, integrating and differentiating actions. The parts Z_0, A, Z_a, Z_b, I, Z_c and D form the model of a continuously acting controller. The square-wave generator G and the oscilloscope *Osc* can be connected at any points.

[1]) J. M. L. Janssen and L. Ensing, The electro-analogue, an apparatus for studying regulating systems, I. Components and functions, Philips Techn. Rev. **12**, p. 257-271, 1951 (No. 9), hereinafter referred to as article I.

In so far as they are of importance, and as far as space permits, details of these parts will be discussed below, but first there are some general questions to be considered.

Electrical "calculating networks" with which mathematical operations like adding, differentiating, integrating, etc. can be carried out, have already been rather extensively described in the literature [2]). For the details we may refer to this literature, but nevertheless a discussion of the essence of the various circuits is not superfluous here.

In the construction of the electro-analogue — an instrument for our own use — the aim has been rather to save time in its development than to economize on material. The number of valves, for instance, — about 350 in all, of which more than half are used in supply units! — could well have been reduced, but the drawback of this large number weighed less than the advantage of the greater chance that the solutions chosen would lead quickly to the object in view.

The next question has greatly influenced the general "plan" of the electro-analogue. This has been so arranged that the input and output terminals of all parts are accessible, so that with the aid of separate leads any desired combination can be obtained. For this purpose the terminals that are to be interconnected must have the same D.C. voltage level, e.g. earth potential. The method of attaining this — with separating capacitors and leak resistors — cannot be considered here on account of the large time constants that would then be needed and which would require impracticably large capacitances.

There would be so many of these *R-C* couplings in the closed control circuit that at a low frequency this circuit would tend to oscillate if the time constants were not chosen carefully. With decreasing frequency each *R-C* coupling gives a greater attenuation and at the same time a phase shift, which ultimately approaches 90 degrees. Oscillation sets in when there is a frequency for which the total phase shift of all *R-C* couplings together is 180° and at the same time the gain of the closed circuit exceeds unity. This state of affairs could be avoided by providing for a sufficient attenuation in one of the couplings to take place at the frequency at which the other couplings begin to contribute appreciably towards the phase shift. Then one *R-C* coupling would have to have a much smaller (but still considerable) time constant than all the others, for which the time constant would have to be very large. Even when the resistances of these couplings are chosen as high as possible, the required capacitances would still have to be impracticably large.

Moreover, the image on the oscilloscope, which after a switching operation becomes displaced (sometimes even disappearing entirely from the screen), would return to its original position only very slowly indeed.

In order to avoid these difficulties, in the parts listed under (1) to (5) the use of blocking capacitors has been completely avoided. In other words, the circuits have been so arranged that in addition to alternating voltages they can also transmit direct voltages.

Many of the parts referred to contain several amplifying valves. Special measures are necessary to preserve the direct-voltage adjustment of these valves and the uniformity in voltage level of the terminals that are to be interconnected. This will be referred to later.

The "special" process analogues

As mentioned in article I, the electro-analogue has thre networks which can serve as process analogues, viz. a universal process analogue with which any step-function response can be obtained, and two "special" networks (*R-C* circuits) with which a limited number of step-function responses can be obtained. Contrary to the case with the universal process analogue, the working of the special process analogues imitates more or less the physical action of the (actual or hypothetical) processes which these analogues represent.

One of the special process analogues consists of twenty *R-C* sections connected direct in cascade, while the other consists of 40 *R-C* sections with intermediate valves preventing the feedback of one stage on the preceding sections. All sections have the same *R* and *C* values. The calculated step-function responses of the last-mentioned network for some values of the number *n* of sections used are given in *fig. 2*.

The two networks are combined electrically and mechanically into one unit. *Fig. 3* represents the circuit. The valves are connected as cathode-followers. This has the advantage that the output impedance is low (about 1000 ohms), so that other parts of the electro-analogue having an input impedance of the order of 1 megohm can be connected to the output without thereby appreciably changing the output voltage. The gain per stage is somewhat less than 1 (about 0.94), but this is no objection here

[2]) See, e.g., F. C. Williams, F. J. U. Ritson and T. Kilburn, Automatic strobes and recurring selectors, J. Inst. El. Engrs 93 IIIA, 1275-1300, 1946, and B. Chance and others, Waveforms (Massachusetts Institute of Technology, Radiation Laboratory Series No. 19), New York 1949.

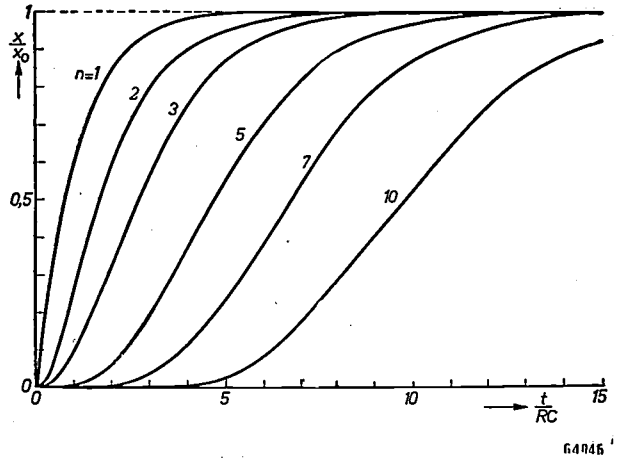


Fig. 2. Step-function response of an R-C network consisting of n sections with amplifying valves connected between the sections, for different values of n.

because the total voltage drop in the process analogue is compensated in the collecting stage.

As mentioned in the introduction, no grid capacitors are used. Consequently the supply has to come from two voltage sources, one supplying about + 200 V and the other —200 V with respect to earth. Each of the cathode resistors ($R_1 + R_2$) has a tapping to which an output terminal is connected and also the resistor R of the following section. The cathode resistors have been so dimensioned that the tappings are at approximately earth potential, while the necessary positive voltage is applied to the cathodes. The potential of the output terminal that is to be used is previously made exactly zero by adjusting the positive supply voltage, which for this purpose can be varied between 160 and 240 V.

Fig. 4 shows the construction in the form of a shelf which can be slid into one of the panels of the electro-analogue. The 40 triode systems are contained in twenty ECC 40 valves, ten of which can be seen on the photograph. On the left is the front of the shelf fully occupied with plug sockets for bringing into action the desired number of sections of the kind required (with or without intermediate valves).

Of course the number of special networks need not be restricted to these two. If necessary the shelf illustrated here can be replaced by one containing some other network required to serve as process analogue.

The "universal" process analogue

For the principle of the universal process analogue reference is made to figures 6, 7, 8 and 9 of article I. The main components are: (1) the delay network, (2) the converting network, and (3) the system with which fractions of voltages derived from the parts sub (1) and (2) can be adjusted and added together. These parts will now be discussed in succession.

The delay network

The delay network consists of 50 sections connected in cascade and built up from inductors and capacitors. These sections have to be so arranged that a voltage step (or any other voltage variation) at the input is transmitted unattenuated and undistorted but with a certain delay τ per section. This means, in the first place, that the network has

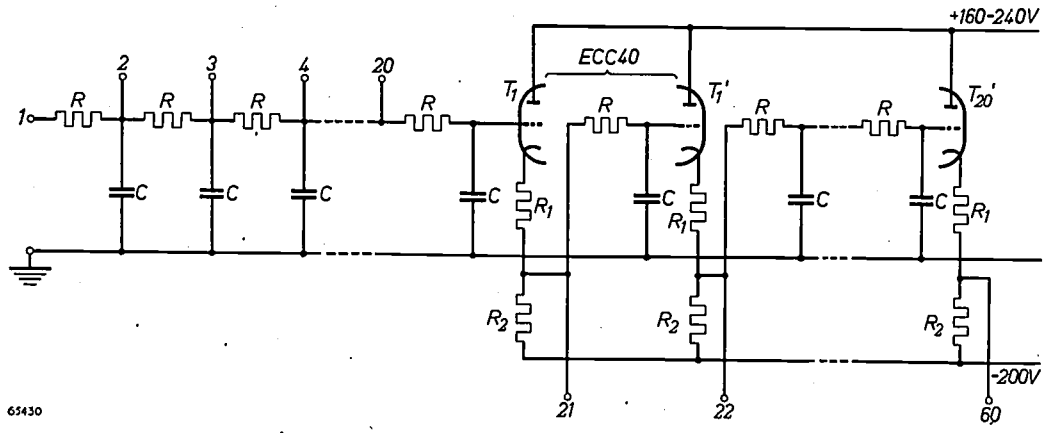


Fig. 3. Twenty R-C sections directly connected in cascade and followed by 40 R-C sections "decoupled" by valves. The valves (double triodes, ECC 40) are connected as cathode followers. The first, T_1, acts as output valve of the network with direct cascade connection, of which terminal *21* is thus always the output terminal. If, for instance, it is desired to use 5 sections of this network, terminal *16* is chosen as input terminal. For the network with interposed valves terminal *20* is always the input terminal; as output terminal one of the terminals *21* . . . *60* is chosen according to the number of sections required.

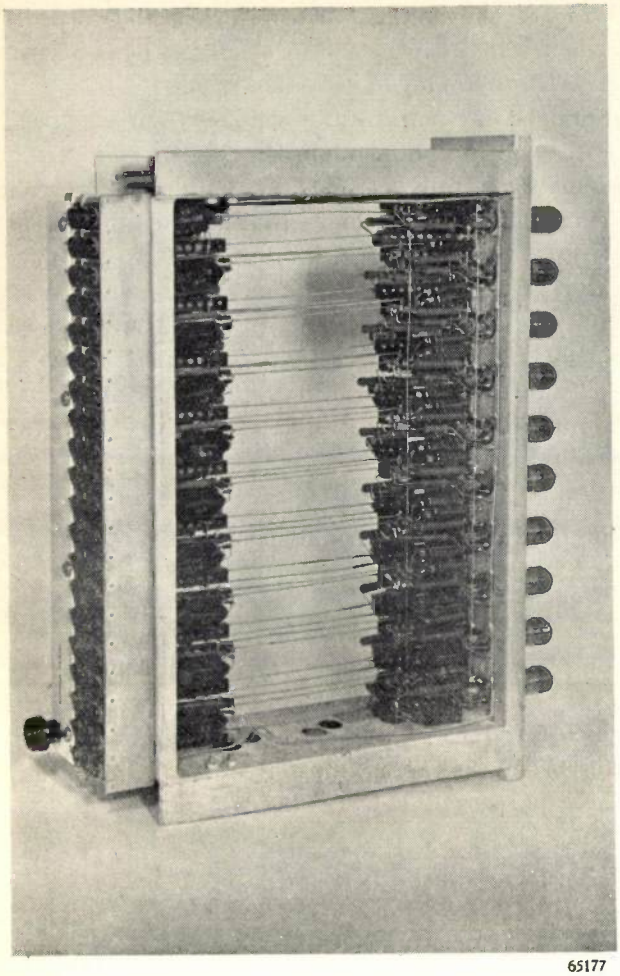


Fig. 4. Shelf containing the two special process analogues according to fig. 3. On the left the plug sockets (*1 ... 60* in fig. 3), on the right ten of the twenty ECC 40 valves can be seen.

to be terminated with its characteristic impedance, so that no reflections occur, and further that each section has to pass on a sinusoidal voltage with a phase shift β proportional to the frequency ($\tau = \beta/\omega$, where ω is the angular frequency).

The circuit of one section is given in *fig. 5a*. It is equivalent to the circuit in fig. 5*b* if $2C_1 = C'$, $L = L' + M'$, $\frac{1}{2}M = M'$ and $\frac{1}{2}C_2 = C''$. The latter circuit has been so drawn that it can be split into two symmetrical parts, one of which is represented in fig. 5*c*. We shall now consider what conditions the quantities have to fulfil in order to give the network the desired properties as near as possible.

The transmission of a section, i.e. the ratio of the output voltage V_2 to the input voltage V_1 (fig. 5*a*), regarding both as complex quantities, can be written as:

$$\frac{V_2}{V_1} = e^{-\Theta} = e^{-(\alpha+j\beta)}, \quad \ldots \quad (1)$$

where the real part α of the complex quantity Θ is a measure of the attenuation taking place in that section. The parameters of the network have to be so chosen that α is zero for all frequencies and β is proportional to ω.

The first condition can be fulfilled exactly (at least in so far as the losses and undesired capacitances of the circuit elements can be ignored by ensuring that

$$\frac{C'}{4C''} = \frac{L' + M'}{L' - M'}, \quad \ldots \quad (2)$$

for which we write m^2.

This formula follows from the condition that the characteristic impedance of the network must be real for all frequencies, which means to say that the product of the no-load and short-circuit impedances of the half section (fig. 5*c*) must have a positive value for all frequencies.

For the further working out of this and the following formulae the reader is referred to the literature on the subject [3]).

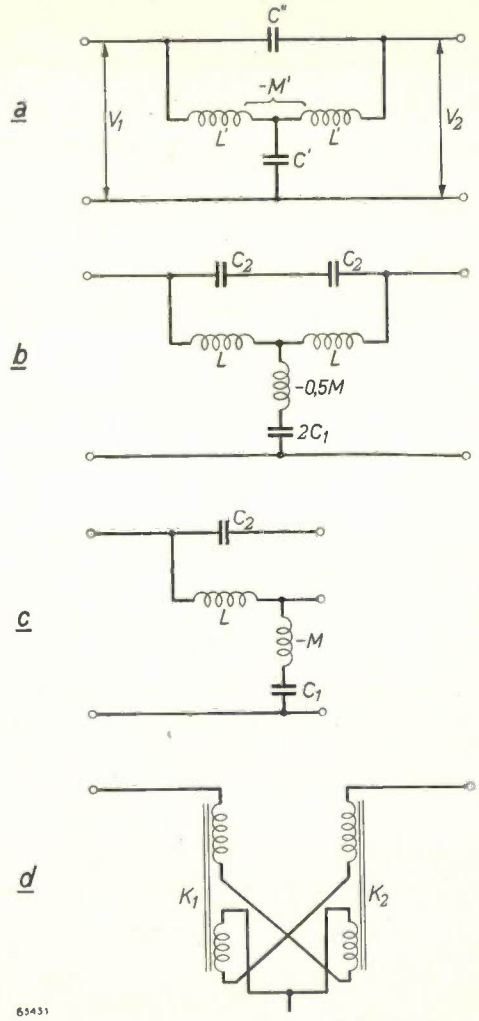


Fig. 5. *a*) Section of the delay network. *b*) Equivalent circuit of (*a*). *c*) Half of the symmetrical circuit (*b*). *d*) Execution of the two coils with self-inductance L' and mutual inductance M'. K_1 and K_2 are separate cores as used in loading coils.

[3]) M. H. Hebb, C. W. Horton and F. B. Jones, On the design of networks for constant time delay, J. appl. Phys. **20**, 616-620, 1949 (No. 6).

The second condition can only approximately be fulfilled. In the angular-frequency range extending from zero to $\omega_0 = 1/\sqrt{LC_2}$ the approximation is as near as possible at a value of m amounting to about 1.6 (see *fig. 6*): in this range τ does not then deviate more than a few per cent from the value τ_0 applying for $\omega = 0$.

In a square-wave voltage all multiples of the fundamental frequency are represented. When such a voltage is applied to the input of the properly terminated delay network which answers the condition (2), with $m = 1.6$, and if $\omega_0/2\pi$ is much greater than the fundamental frequency of the applied voltage, then the fundamental frequency and its multiples up to $\omega_0/2\pi$ are transmitted with a practically equal delay $\tau \approx \tau_0$ per section, whilst the higher harmonics are transmitted with less delay (fig. 6). Hence the voltage at the output of, say, the tenth section (if the network is properly terminated there) appears as in *fig. 7a*: after $t = 10\tau$ there is in fact a fairly good reproduction of the applied voltage step, but prior to $t = 10\tau$ there is an oscillation with gradually diminishing frequency and increasing amplitude [4]). In this case $\tau = 56.3$ μsec, $\omega_0/2\pi = 9120$ c/s and the fundamental frequency of the square-wave voltage is 50 c/s.

If, on the other hand, the applied voltage is more or less trapezoidal, with flanks covering a time

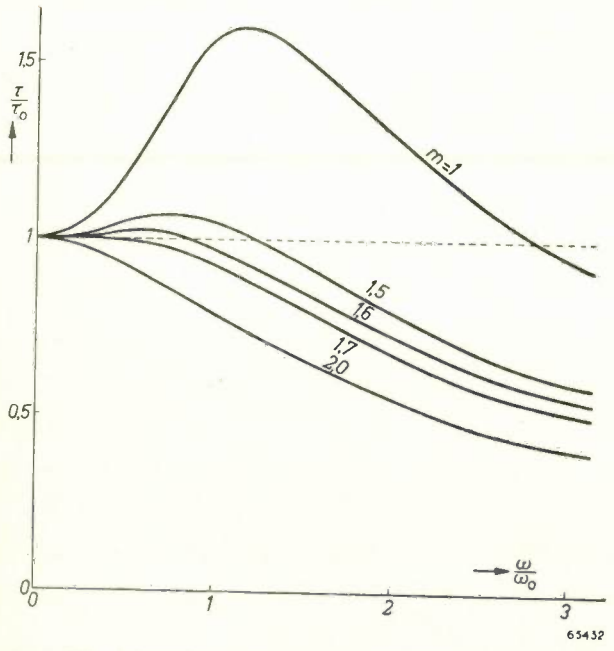
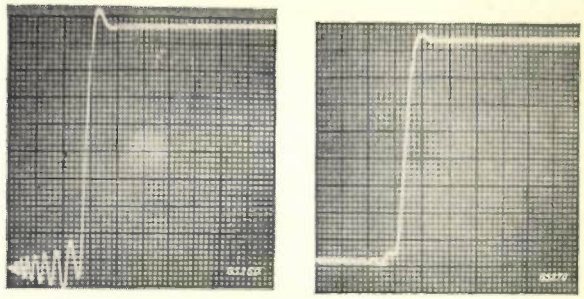


Fig. 6. The delay time τ per section according to fig. 5 (when eq. (2) is fulfilled) as a function of the angular frequency ω for different values of the parameter m. (For the meaning of τ_0, ω_0 and m see the text.) For $m \approx 1.6$, τ is practically constant in the range $0 \leqq \omega \leqq \omega_0$.



a *b*

Fig. 7. Oscillograms of the voltage at the output of the tenth section of the delay network when this is terminated immediately behind that section. In (*a*) at the input of the network a voltage step was applied, and in (*b*) a voltage gradually rising from zero to maximum in about the time τ and then remaining constant. (Owing to the constricted time scale it is not very clearly shown that in (*b*) the steepest part of the curve is much less steep than in (*a*).)

about as long as the delay τ of the network, then the harmonics with frequencies higher than $\omega_0/2\pi$ are so weak that the oscillation is invisible (fig. 7*b*). This is one of the reasons why the square-wave voltage is in fact converted into a practically trapezoidal voltage. The other reason is that the output voltage of the process analogue then has the shape of a series of straight sections much more closely resembling a continuous curve than the echelon-shaped curve which would be obtained with a square-wave input voltage (cf. article I, figs 7 and 9). Owing to various causes the curve actually obtained is in fact continuous.

The converting network

The square-wave generator voltage is converted into a trapezoidal voltage in a network (*0, fig. 8*) for which ten identical *R-C* sections with intermediate valves have been chosen. Thus voltages are obtained having the shape as represented in fig. 2. The voltage for $n = 10$ is suitable for our purpose provided the time constant *RC* is suitably chosen, such that the "dead time" between the voltage step at the input and the appearance of the output voltage is practically equal to the delay time τ of the filter. If it is desired to imitate a process which on the time scale of the analogue has a shorter "dead time", or none at all, then this can be done

[4]) The question arises why, instead of a network uniformly transmitting voltages of all frequencies, use is not made of a low-pass filter with a cut-off frequency round about $\omega_0/2\pi$, for then the higher harmonics in question would be suppressed. The answer to this is that, with a voltage step on the input, a low-pass filter always shows a natural oscillation, which in the present case is likewise very disturbing, no matter how well the filter is terminated (see, e.g., Philips Techn. Rev. **12**, 239, 1951 (No. 8), fig. 11*b*).

by including in the output voltage of the process analogue also other voltages from intermediate sections of the converting network, say from the first, the third and the sixth sections (see fig. 8).

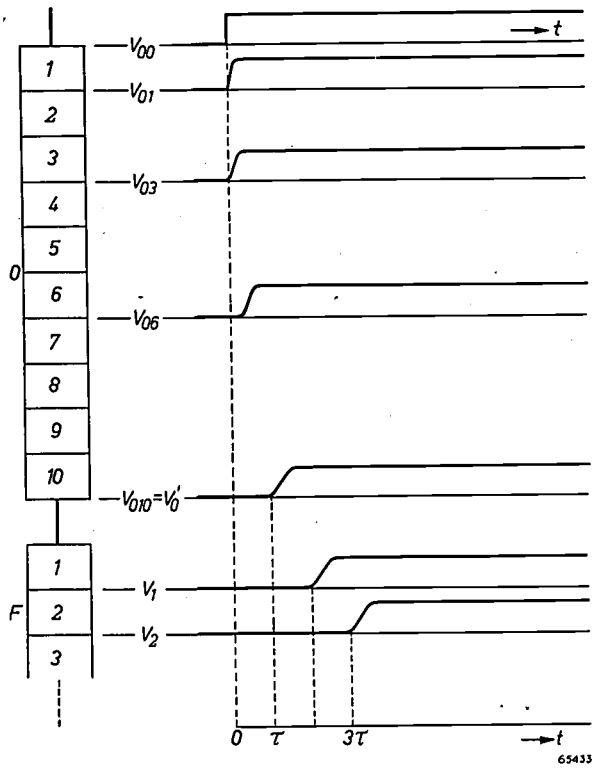


Fig. 8. Delay network F preceded by a network 0 converting the voltage step V_{00} into the gradually rising voltage V_{010} (referred to in article I as V_0').

For the details of the circuit see the right-hand half of fig. 3.

Varying and adding fractions of the output voltages of the delay and converting networks

For the process analogue to give the conditions of any arbitrary process, i.e. for adjusting any arbitrary step-function response curve, variable fractions of the voltages $V_1 \ldots V_{50}$ at the outputs of the sections of the delay network have to be added together; it may be necessary to add also fractions of the voltages V_{01}, V_{03}, V_{06} and V_{010} from the converting network. The result is a sum voltage V_s':

$$V_s' = [\alpha_{01} V_{01} + \alpha_{03} V_{03} + \alpha_{06} V_{06} + \alpha_{010} V_{010}] + \\ + [\alpha_1 V_1 + \ldots + \alpha_{10} V_{10}] + [\alpha_{11} V_{11} + \ldots + \alpha_{20} V_{20}] + \\ + \ldots + [\alpha_{41} V_{41} + \ldots + \alpha_{50} V_{50}] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

V_s' may show the shape of any desired step-response curve, provided the coefficients α in (3) can be varied independently of each other from $-\alpha_{max}$ to $+\alpha_{max}$. This variability is obtained by amplifying each of the 54 voltages $V_{01} \ldots V_{50}$ in a push-pull stage as shown in *fig. 9* [5]). The output voltage V_0 appears at the sliding contact of a potentiometer R_4 between the anodes and by means of this contact can be varied from $-\frac{1}{2}\mu$ to $+\frac{1}{2}\mu$ times the input voltage V_i; here μ represents the total gain of the stage.

For adding up the output voltages the 54 push-pull amplifiers are divided into six groups, denoted in (3) by square brackets. The first group has four amplifiers, while the other five groups, connected to the sections of the delay network, each have ten.

Adding is done in two steps: first the four, or ten, output voltages from the push-pull amplifiers belonging to one group are added together in a "sub-collecting stage", after which the six group voltages obtained in this way are added together, this taking place in the output collecting stage of the process analogue (S_2 in fig. 1).

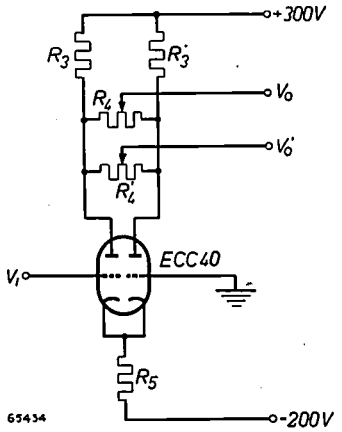


Fig. 9. Push-pull circuit for amplifying the output voltages of the converting network and of the delay network. R_3, R_3' anode resistors of the double triode ECC 40. The high cathode resistance R_5 ensures symmetry in spite of the asymmetrical input. With the potentiometer R_4 the respective coefficient α of eq. (3) can be varied from $-\alpha_{max}$ to $+\alpha_{max}$. In each push-pull amplifier there is a second potentiometer (R_4'), independent of the first, for adjusting a second step-response curve.

Fig. 10 shows the circuit of one of the sub-collecting stages [6]). By means of equal resistors R_6 the output terminals of the push-pull amplifiers of one group are connected to one collecting point (in this case G_2). The voltage obtained at this point is the average of the voltages at the output terminals and thus, except for a constant factor (4 or 10),

[5]) The working of this push-pull circuit, which has an asymmetrical input, is explained, i.a., in an article by E. E. Carpentier in Philips Techn. Rev. 9, 204, 1947, fig. 3b.

[6]) Similar (and other) adding networks are described in the book (pages 629-648) quoted in footnote [2]).

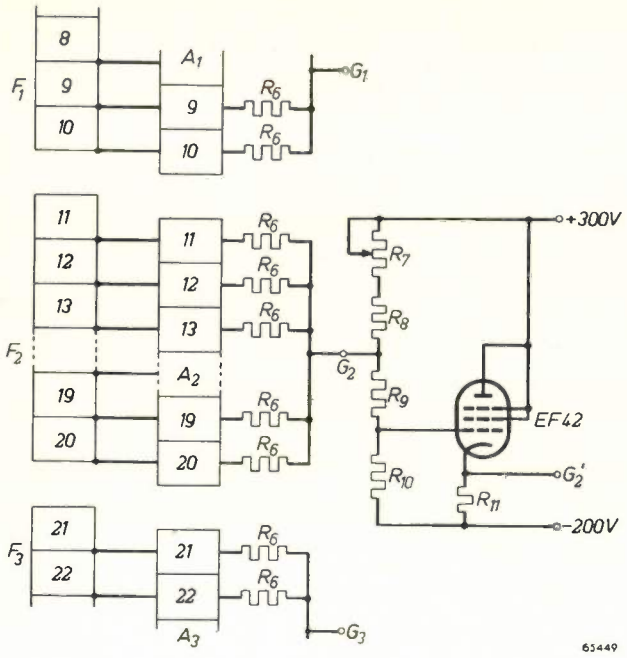


Fig. 10. Sub-collecting stage, in which the ten output voltages of a part (F_2) of the delay network are added together after amplification by the push-pull amplifiers A_2 (fig. 9). (F_1 the preceding, F_3 the following ten sections, with corresponding amplifiers A_1 and A_3.) The outputs of the amplifiers are connected via resistors R_6 to collecting points G_1, G_2, G_3, etc. The EF 42 valve, used as a cathode follower, and the resistors R_7 ... R_{11} provide for a low-impedance output with a zero direct-voltage level.

is equal to the desired sum. Via a cathode follower (this is necessary owing to the low output impedance needed) the voltage is transmitted from G_2 to the group output terminal G_2'. By means of a potentiometer formed by the resistors R_7 ... R_{10}, with R_7 variable, the direct voltage at G_2' is brought down to zero (that at G_2 is about + 150 V). This potentiometer is so dimensioned that in the state of rest no current flows through the ten resistors R_6 and the sliding contacts of the potentiometers R_4 (fig. 9), no matter in what position these sliding contacts may be; thus the direct-voltage level of the collecting points does not change when the coefficients a are varied.

The second step in the addition is the adding together of the six group output voltages. As already stated, this takes place in the output collecting stage of the process analogue (S_2 in fig. 1). This stage also affords an opportunity to add to the sum a sinusoidal or step "disturbance", so that it can be ascertained how the controlling circuit reacts to such a disturbance (examples of this have been given in article I). The construction of this collecting stage will be described below.

Construction of the universal process analogue

The universal process analogue is contained in six shelves, each carrying one of the groups previously mentioned. Such a shelf is illustrated in *fig. 11*. In the middle there are 20 coils, pairs of which belong to each section of the delay network (fig. 5*a*). The necessary mutual inductance M' is obtained by arranging the coils as represented in fig. 5*d*. At the top on the right-hand side of the shelf shown in fig. 11 is the valve EF 42 of the sub-collecting stage (cf. fig. 10), and underneath that are the ten double triodes ECC 40 of the push-pull amplifiers (fig. 9). As indicated in the last-mentioned illustration, in addition to the potentiometer R_4 there is another, R_4'; accordingly also the sub-collecting stage is duplicated. The reason for this is that by this means two different step-responses can be obtained, either for two values of one of the working quantities (in the case of temperature control, for instance, for two ambient temperatures), or for two detecting elements at different points in the process (one serving for the control and the other as a means of observing the behaviour of the controlled quantity elsewhere in the process). On the left-hand side of the shelf in fig. 11 are the 20 knobs of the potentiometers R_4 and R_4'.

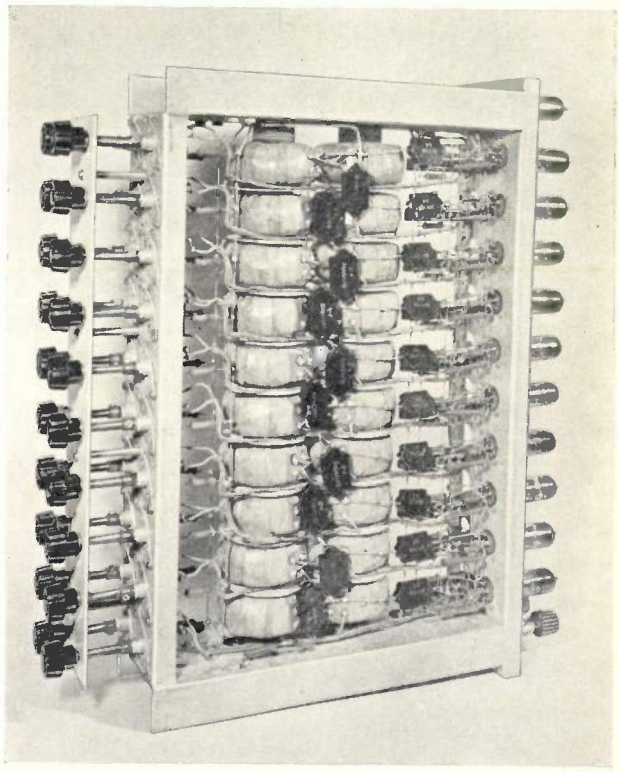


Fig. 11. Shelf with ten sections of the delay network. On the left the knobs of the 20 potentiometers R_4 and R_4' (fig. 9), in the middle the coils and capacitors (fig. 5), and on the right the amplifying valves.

Fig. 12 shows the front of the electro-analogue. In the panel on the right are five of the shelves (*F*) described. Underneath these is a shelf *Q* containing the two special process analogues previously described. If it is desired to work with the universal process analogue the bottom shelf has to be replaced by another containing the converting network with accessories.

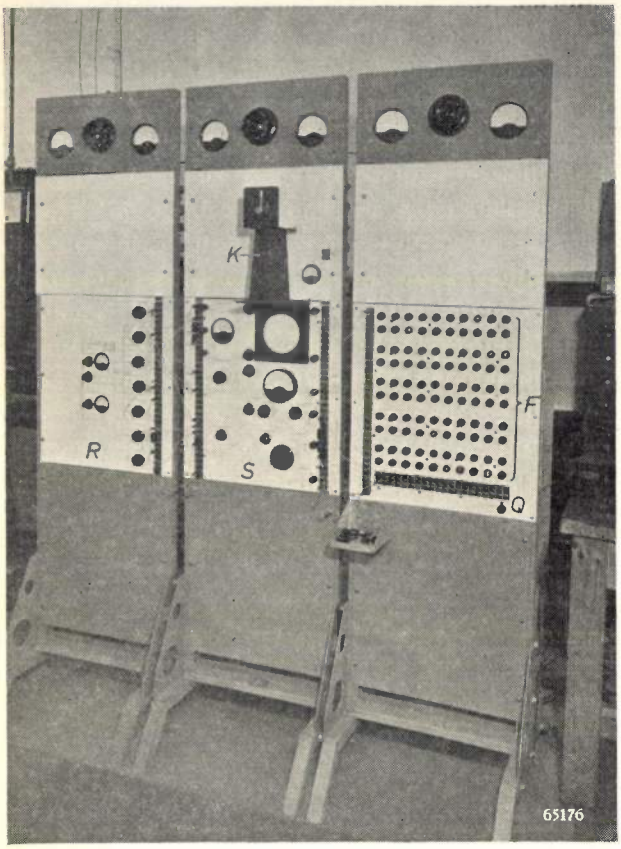


Fig. 12. Front of the electro-analogue. *R* panel with controller models, *S* panel with oscilloscope, collecting stages, square-wave generator and *R-C* generator, *F* delay network, *Q* special process analogues. When the universal process analogue is used *Q* is replaced by a shelf with the converting network (*O* in fig. 8). *K* cylinder which can be turned down for photographing the oscillograms.

The "universal" amplifier

Amplifiers are used at several points in the controlling circuit of the electro-analogue, not only at the point marked *A* in the block diagram (fig. 1) but also in the collecting stages (S_1 and S_2), in the integrator (*I*), in the differentiator (*D*) and in the oscilloscope (*Osc*). Apart from small modifications, most of these amplifiers are built according to the same principle, and that is why we speak here of an "universal" amplifier.

Briefly, the universal amplifier consists of three push-pull stages in cascade, the first with an asymmetrical input and the last one connected as a cathode follower in order to get a low output impedance. As already said, there are no grid capacitors, because these would lead to phase shifts and thus cause the controlling circuit to oscillate. By applying negative feedback a high degree of stability is obtained.

For a D.C. voltage amplifier a push-pull circuit offers several advantages. When a pair of valves connected in push-pull have a common cathode resistor and a common screen-grid resistor, provided there is absolute symmetry, the currents through these resistors are constant, regardless of the input voltage, and thus also the cathode voltage and the screen-grid voltage are constant. These resistors, therefore, do not involve any reduction of the gain (as they would do in an asymmetrical amplifier unless they are shunted by large capacitors, a remedy which of course is ineffective for a direct voltage). In the second place, any differences arising between a pair of valves only slightly disturbs the symmetry: the common resistors aim, as it were, at preserving the symmetry (see footnote [5])). Finally, in the electro-analogue it is convenient to have a symmetrical output voltage even if only one of the two voltages (with respect to earth) is used: for the control one can take the voltage that has the right sign; without such a possibility it would in some cases be necessary to have an "inverter stage" for reversing the sign of the voltage.

The measures described are applied in the somewhat simplified circuit of the universal amplifier (*fig. 13*), which is explained in the subscript.

In order to avoid any disproportionality in, for instance, the push-pull amplifiers (fig. 9) behind the delay network, once the D.C. voltage levels have been adjusted they must not be liable to any change. It is therefore necessary that the supply voltages for the amplifier are absolutely constant. The anode and grid voltages are obtained from carefully stabilised rectifiers (+600 V, +300 V and −300 V). A fourth stabilised rectifier supplies the filament current for the six pentodes in the amplifier: in the case of A.C. supply from the mains the mains fluctuations would cause too great asymmetries — in spite of the aforementioned "aim at symmetry" — owing to the two valves in a push-pull stage never being exactly identical. Since D.C. supply is used, valves of the U type have been chosen (UF 41) with a filament current of only 100 mA at 12.6 V. Eighteen filaments connected in series are fed from one rectifier.

The over-all gain (the ratio of the output voltage, with respect to earth, to the input voltage) is 200 (46 dB).

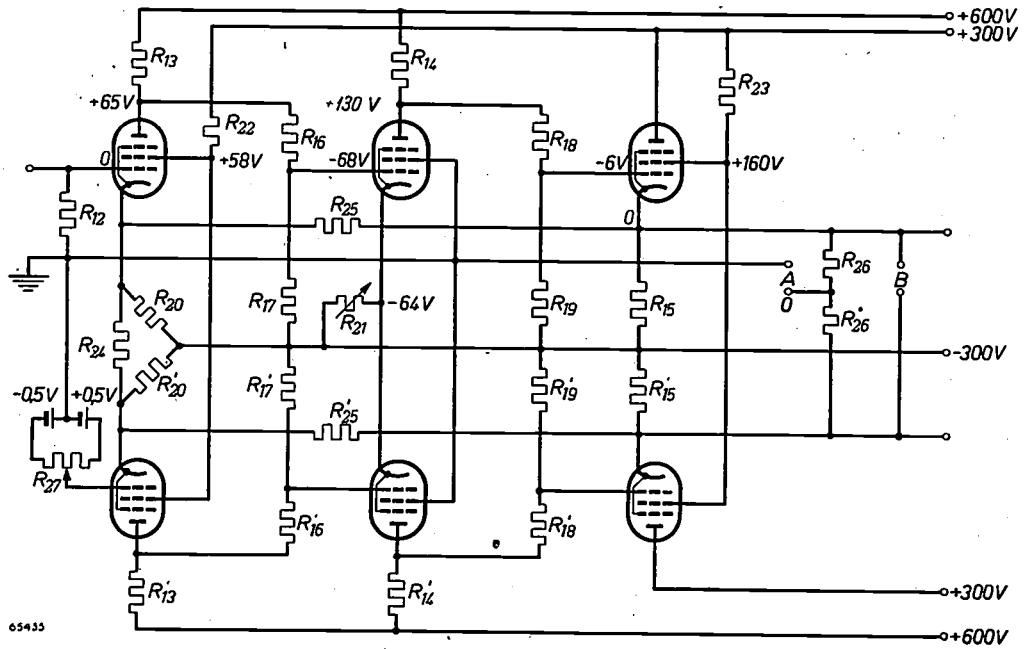


Fig. 13. Circuit diagram of the universal amplifier (simplified). R_{12} input resistor, R_{13}, R_{13}' anode resistors of the first stage, R_{14}, R_{14}' ditto of the second stage, R_{15}, R_{15}' cathode resistors of the third stage (cathode-follower circuit). R_{16}-R_{17}, R_{16}'-R_{17}', R_{18}-R_{19}, R_{18}'-R_{19}' form voltage dividers by means of which the control grids of the next stage are brought to the right potential. R_{20}-R_{20}' (to be imagined, to a first approximation, as being connected in parallel, since R_{24} has a much smaller resistance) is the common cathode resistor of the firsts tage, R_{21} is that of the second stage. R_{22} and R_{23} are the common screen-grid resistors of the first and third stages respectively. R_{24} is a resistor for the negative feed-back from the output terminals via the equal resistors R_{25} and R_{25}'.

For checking the direct-voltage level and the symmetry the input is short-circuited and a moving-coil voltmeter is connected in succession to A and to B. At A the mean value of the voltage at the output terminals ($R_{26} = R_{26}'$) is measured with respect to earth; this voltage is reduced to zero by adjusting R_{21}, which affects in equal sense the cathode currents and thus also the anode currents and anode voltages of the second stage. At B the potential difference between the two output terminals is measured, and thus the degree of asymmetry; this difference is brought to zero with the aid of the potentiometer R_{27}, with which the control-grid voltage of one of the valves in the first stage can be varied from about −0.5 V to +0.5 V (actually R_{27} is not fed from a battery but from the +300 V and −300 V sources).

The correct D.C. voltages (with respect to earth) of the various points are indicated in the diagram.

The six valves are of the UF 41 type.

The collecting stages

The purpose of the collecting stage S_1 (fig. 1) is to form the sum of the output voltages from the parts comprising the continuously acting controller, to which sum it should be possible to add a disturbing voltage. In the collecting stage S_2 — when the universal process analogue is used — it is the output voltages from the sub-collecting stages that have to be added together, possibly also with a disturbing voltage. Furthermore, the collecting stages together have to provide for a gain which at least compensates the attenuation taking place in the process analogue. As mentioned in the discussion of the oscillograms in article I, what is desired is that a step voltage with amplitude s applied to the input of the process analogue (after this has temporarily been disconnected from the controller) should produce at S_2 an output voltage with final value x_∞ exactly equal to s, because then it is easiest to interpret the oscillograms. Now the fact that the collecting stages also have to function as amplifiers makes it impossible for them to have the same circuit as the sub-collecting stages (fig. 10), for the "gain" of the latter is less than unity. Use is therefore made of a "universal" amplifier (fig. 13), but with a small modification related to the following question.

In the sub-collecting stages the adding together of n voltages ($V_1 \ldots V_n$) is done with the aid of n equal resistors (R_6, fig. 10) coming together at one collecting point. This point assumes a potential equal to the mean value of the n voltages $(V_1 + \ldots + V_n)/n$. If n is constant, as it is in the case of the

sub-collecting stages (namely either always 10 or always 4), then this mean value differs from the desired sum, $V_1 + \ldots + V_n$, only by a constant factor $1/n$. In the case of the collecting stages that we are now about to consider, however, n is variable: for S_1 it varies from $n = 1$ (in the case of a controller with, for instance, only a proportional term) to $n = 6$ (controller with a proportional, a single and a double integrating, and a single and a double differentiating term, plus a disturbance voltage), for S_2 from $n = 1$ (when one of the special process analogues is employed) to $n = 7$ (six sub-collecting stages of the universal process analogue plus a disturbance voltage). The effect of a variation of n could be compensated with a corresponding but opposite variation of the gain, but a simpler method is that described below whereby the sum of the voltages is obtained with a proportionality factor practically independent of n.

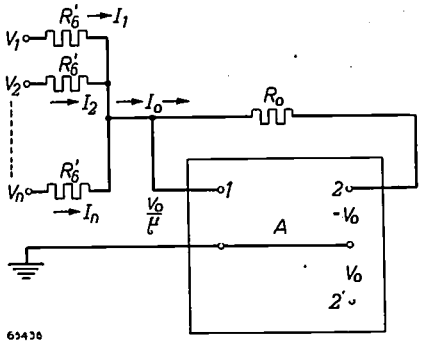


Fig. 14. Circuit diagram of a collecting stage. $V_1 \ldots V_n$ are the voltages (variable in number) that are to be added together. A universal amplifier according to fig. 13. When the gain μ is high the output voltage is approximately equal to the product of the sum $(V_1 + \ldots + V_n)$ and the ratio of the resistances R_0 and R_6'.

In *fig. 14* the amplifier A has an asymmetrical input and a symmetrical output, as, for example, the universal amplifier previously discussed. The input terminal *1* is connected via equal resistors R_6' to the n points from which the voltages $V_1 \ldots V_n$ have to be added together, and, moreover, via a resistor R_0 to the output terminal *2* where the voltage is in counter-phase to that at terminal *1*. The output voltage (with respect to earth) is denoted by V_o. If the gain μ is very high, then to a first approximation the voltage V_o/μ at terminal *1* is negligible compared with the input voltages $V_1 \ldots V_n$, so that for the currents $I_1 \ldots I_n$ flowing through the resistors R_6' we then have:

$$I_1 \approx \frac{V_1}{R_6'}, \ldots, I_n \approx \frac{V_n}{R_6'}. \quad \ldots \quad (4)$$

If V_o/μ is also negligible with respect to V_o then the current I_0 flowing through R_0 is:

$$I_0 \approx \frac{V_o}{R_0} \quad \ldots \ldots \ldots \quad (5)$$

From:

$$I_0 = I_1 + I_2 + \ldots + I_n$$

it follows, by applying the equations (4) and (5), that

$$V_o \approx \frac{R_0}{R_6'}(V_1 + \ldots + V_n).$$

From this formula it is seen that — provided the gain μ is high enough — the output voltage is equal to a fixed ratio (independent of n) multiplied by the sum of $V_1 + \ldots + V_n$.

From an exact calculation, however, we find:

$$V_o\left(1 + \frac{1}{\mu} + \frac{nR_0}{\mu R_6'}\right) = \frac{R_0}{R_6'}(V_1 + \ldots + V_n),$$

in which the quantity n still occurs. With the values used for R_0 (about 5 megohms) and R_6' (1 megohm) and with the normal gain of the universal amplifier ($\mu = 200$) the coefficient of V_o is equal to 1.03 for $n = 1$ and 1.18 for $n = 7$, so that there is still rather a great dependency upon n. For this reason, in the amplifiers employed in the collecting stages the resistor R_{24} (fig. 13) for the negative feedback is short-circuited, so that μ is about 1500 and the coefficient of V_o varies only from 1.004 to 1.024.

There is no objection against the negative feedback being cut out by shorting R_{24}, since the resistor R_0 introduces a new and much heavier feedback. By making R_0 variable (from 0.85 to 5.4 megohms) the gain of the collecting stage can be varied within certain limits (in the present case from 0.85 to 5.4).

The integrator and the differentiator

As far as the circuit is concerned, the integrator and the differentiator are only modified forms of the collecting stage. When, in the diagram of the latter (fig. 14), we take only one of the n input resistors (now to be called R) and replace R_0 by a capacitor C, then, as will be seen from what follows, we have an integrator (*fig. 15*), and when interchanging R and C we have a differentiator.

---

[7]) For differentiating and integrating networks see, e.g., pages 648-666 of the book quoted in footnote [2]).

On account of the capacitor C between output and input, an integrator according to the system of fig. 15 is known under the name of a Miller integrator, named after the first scientist to study the influence of capacitance between anode and grid of a triode, though he did not then have an integrating effect in mind (J. M. Miller, Sci. Papers Nat. Bur. Stand. No. 351, 15, 367-385, 1919-1920).

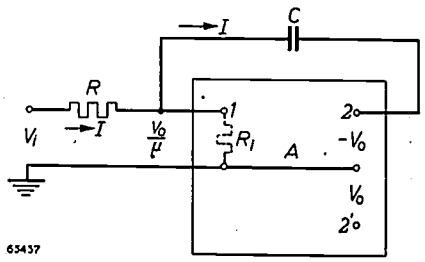


Fig. 15. Basic circuit of the integrator. A is a universal amplifier according to fig. 13.

The integrator

If the universal amplifier A again has a very high gain μ, then to a first approximation the voltage V_o/μ at the input of A is negligible compared with the input voltage V_i. The current I flowing through R and C is then:

$$I = \frac{V_i}{R} = C\frac{dV_o}{dt}, \quad \ldots\ldots \quad (6)$$

or

$$V_o = \frac{1}{RC}\int V_i\, dt,$$

so that — at least when ignoring V_o/μ — the output voltage is proportional to the time integral of the input voltage.

To find the frequency characteristic of the integrator we let V_i change sinusoidally with time, with angular frequency ω. Equation (6) then becomes

$$I = \frac{V_i}{R} = V_o \cdot j\omega C,$$

or

$$\mu' = \left|\frac{V_o}{V_i}\right| = \frac{1}{\omega CR}.$$

Thus the characteristic varies in inverse proportion to the frequency (curve *1* in *fig. 16a*). The finite value of μ, however, prevents an unlimited increase of the quantity μ' with decreasing frequency; the actual frequency characteristic is therefore a curve like *1'* and does not exceed the value μ.

If account is taken of the finite value of μ then account should also be taken of the finite value of the input resistance R_i (indicated by the broken line in fig. 15). This R_i is the resistor R_{12} in fig. 13. According to the valve specifications this resistance must not exceed 1 megohm, which value has been chosen for the resistor R.

Better expression is given to the behaviour of the integrator when the two coordinates of the frequency characteristic are on a logarithmic scale (fig. 16*b*): the hyperbola in fig. 16*a* then becomes a descending straight line *1*, corresponding to the behaviour of an ideal integrator. If the integrator is not an ideal one then with decreasing frequency this straight line ultimately changes into a horizontal line *2* lying at a level determined by the gain μ. The abscissa of the point of intersection of the two straight lines (for the sake of simplicity the curved transition is not usually drawn) indicates roughly the lowest frequency for which the integrator can still be used.

As is immediately seen from fig. 16*b*, the higher the gain μ, the lower is this limit. With the values chosen for R and C (respectively 1 megohm and 1000 pF), upon which, of course, the position of the point of intersection also depends, the value $\mu = 200$ of the universal amplifier gave a frequency limit which was not low enough for our purpose. Just as in the case of the collecting stage, the gain might be increased by short-circuiting the resistor for the negative feedback, but for the integrator

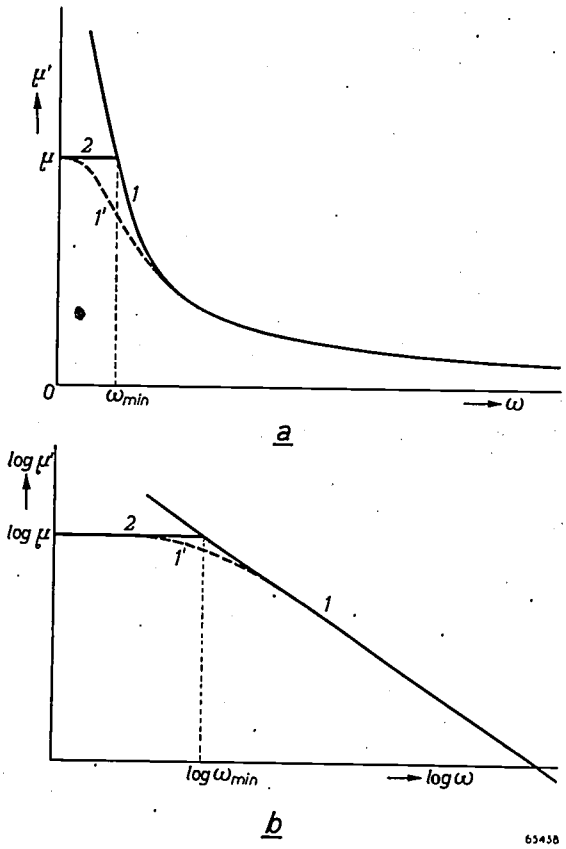


Fig. 16. *a*) Frequency characteristic of the integrator, on a linear scale. Curve *1* applies for an ideal integrator ($\mu = \infty$), curve *1'* for an integrator with finite value of μ. *b*) The same but with both coordinates on a logarithmic scale. The characteristic of the ideal integrator is the straight line *1*, which, when μ is finite, at low frequencies ultimately changes to the horizontal line *2* (see also (*a*)). The continuous transition *1'* between *1* and *2* is not usually drawn. A line dropped from the point of intersection of *1* and *2* to the abscissae indicates the lowest frequency (ω_{min}) for which the integrator can still be used.

this appeared to lead to instability. Therefore another method has been followed, which is explained in the following small-type text and which offers the theoretically interesting possibility of obtaining an ideal integrator.

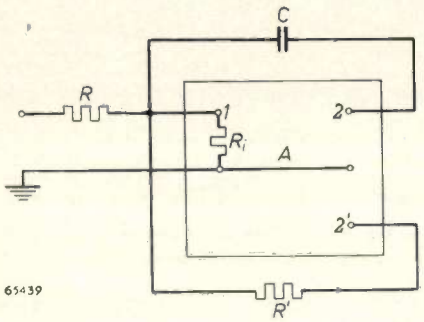


Fig. 17. The working of the integrator can be improved by adding a given resistance R' between the terminals *1* and *2'* of the system represented in fig. 15.

When the input terminal *1* of the amplifier (see *fig. 17*) is connected not only via the capacitor C to the output terminal *2* but also via resistor R' to the other output terminal *2'* (where the voltage is in phase with that at terminal *1*), a simple calculation shows that μ has to be replaced by the apparent gain μ_1;

$$\mu_1 = \frac{\mu}{1 + \frac{R}{R_i} - \frac{R}{R'}(\mu - 1)},$$

where R_i is the input resistance of the amplifier. This expression becomes infinitely large when R' is so chosen that

$$1 + \frac{R}{R_i} - \frac{R}{R'}(\mu - 1) = 0,$$

or

$$R' = (\mu - 1)\frac{RR_i}{R + R_i}.$$

If this condition is fulfilled an ideal integrator is obtained but it is then on the limit of direct-voltage instability. This does no harm so long as the control circuit is closed, but the presetting of the amplifier (resistors R_{21} and R_{27}, see fig. 13) has to be done with the circuit open, and then difficulties arise. For R' we have therefore chosen a somewhat larger value than is indicated by the last formula.

As a matter of fact the question whether the integrator is ideal or not is more of a theoretical than practical importance, since the integrator of the actual controller to be imitated is not ideal either. And the same applies to differentiators.

By way of an example oscillograms of the input and output voltages of the integrator are shown in *fig. 18*.

The differentiator

When R and C in the integrator (fig. 15) are interchanged we have a differentiator (*fig. 19*). This is easily understood when again V_o/μ is considered, to a first approximation, to be negligible compared with V_i and V_o. The current I then flowing through C and R is

$$I = C\frac{dV_i}{dt} = \frac{V_o}{R}, \quad \ldots\ldots \quad (7)$$

or

$$V_o = RC \cdot \frac{dV_i}{dt}.$$

The frequency characteristic of the ideal

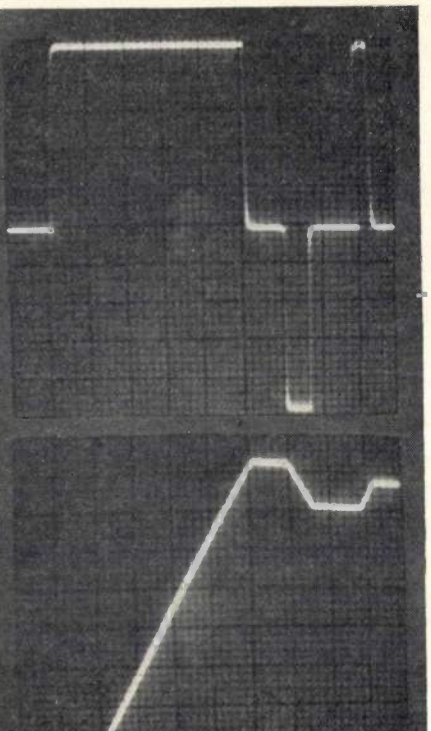


Fig. 18. To illustrate the working of the integrator a voltage of the shape of the oscillogram (*a*) is applied to the input. In (*b*), the oscillogram of the output voltage, the integral of (*a*) is easily recognized.

The input voltage (*a*) was derived from the model of the controller with three-step action, which together with the integrator forms a model of a controller with three speeds (cf. I, fig. 28).

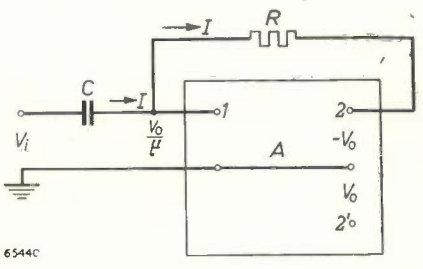


Fig. 19. By interchanging R and C in the integrator (fig. 15) a differentiator is obtained.

differentiator is found by introducing for V_i in (7) a sine function:

$$I = j\omega C\, V_i = \frac{V_o}{R},$$

and hence:

$$\mu' = \left|\frac{V_o}{V_i}\right| = |j\omega CR| = \omega CR.$$

Thus μ' is proportional to the frequency, and $\log \mu'$ increases linearly with $\log \omega$ (line *1* in *fig. 20*).

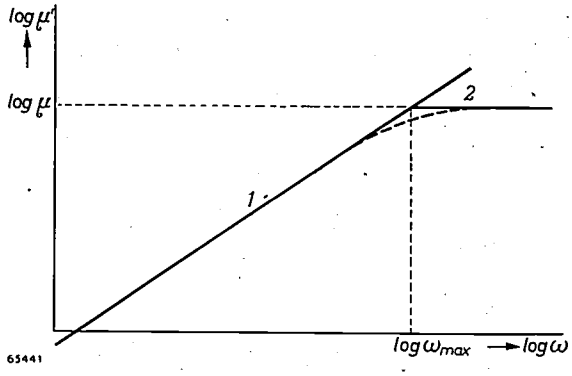


Fig. 20. Frequency characteristic of the differentiator, with both coordinates on a logarithmic scale. In analogy with fig. 16*b* the characteristic is schematically drawn to show two straight lines, *1* (characteristic of the ideal differentiator) and *2*. A line dropped from the point of intersection of *1* and *2* to the abscissae corresponds roughly to the highest frequency ω_{max} at which the differentiator still works satisfactorily.

Owing to the finite value of μ, however, with increasing frequency $\log \mu'$ ultimately runs horizontally (*2* in the same figure) at the level of $\log \mu$. Taking the point of intersection of the two straight lines as indicating the maximum frequency at which the differentiator still works satisfactorily, it is immediately seen that the larger the value of μ the higher is this limit. In the differentiators of the electro-analogue, with $C = 1000$ pF and $R =$ 1 megohm, in the normal state the universal amplifier gives, with $\mu = 200$, a sufficiently high frequency limit.

For double integrations or double differentiations, two integrators or two differentiators, as the case may be, can be connected in cascade.

It is to be pointed out here that a characteristic with a proportional and a number of integrating and differentiating terms is certainly not to be regarded as *the* universal characteristic of a (continuously acting) controller; also other characteristics are possible, even when disregarding non-linear characteristics. For this reason space has been reserved in the electro-analogue for networks built up according to need to be used as a model for a continuously acting controller.

Models of discontinuously acting controllers

A circuit which can serve as model of a controller with two-step action is the flip and flop circuit, i.e. a circuit having two valves (*fig. 21*) only one of which can pass current at a time, the other thereby being cut off. When the conducting valve receives a negative voltage pulse on the grid it is cut off, the other valve then being unblocked and thus becoming conducting. When a negative voltage pulse is applied to the grid of the latter valve this in turn is cut off, and so on. Just as is the case with a controller with two-step action, there are therefore two stable conditions [8]). Moreover, a certain backlash can be obtained in the change-over so that the circuit imitates a controller with backlash (fig. 21*b* in article I), the characteristic of which is similar to a hysteresis loop.

Fig. 22 represents, in a somewhat simplified form, the circuit which in the electro-analogue functions as model of a controller with two-step action. The width of the hysteresis loop can be adjusted according to need. The working of the circuit is explained in the subscript.

Here we can only very briefly consider a controller with three-step action. The principle is illustrated in *fig. 23*. The output terminal *3* has a positive potential when only the valve T is conducting, earth potential so long as both valves are cut off, and a negative potential when only the valve T' is conducting. Thus three stable states are possible. This is the principle applied in the actual circuit, which for the rest resembles very much that of fig. 22.

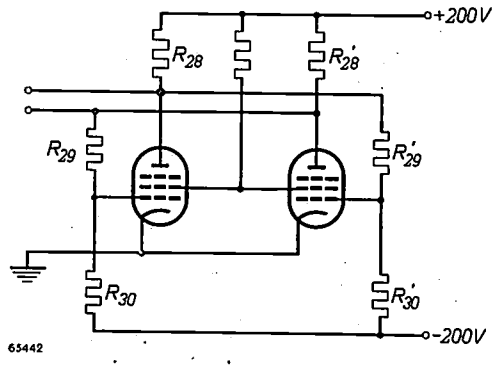


Fig. 21. Flip-and-flop circuit. Only one of the two valves is conducting at a time and thereby cuts off the other. The position is reversed when a negative voltage pulse is applied to the grid of the conducting valve. The anode resistors R_{28}, R_{28}', together with respectively the resistors R_{29}'-R_{30}' and R_{29}-R_{30}, form voltage dividers ensuring the correct adjustment.

[8]) The same system is also applied in an entirely different field, viz. in the "scale of two" circuits of counting apparatus; see, e.g., Philips Techn. Rev. **6**, 76, 1941 and **10**, 7, 1948.

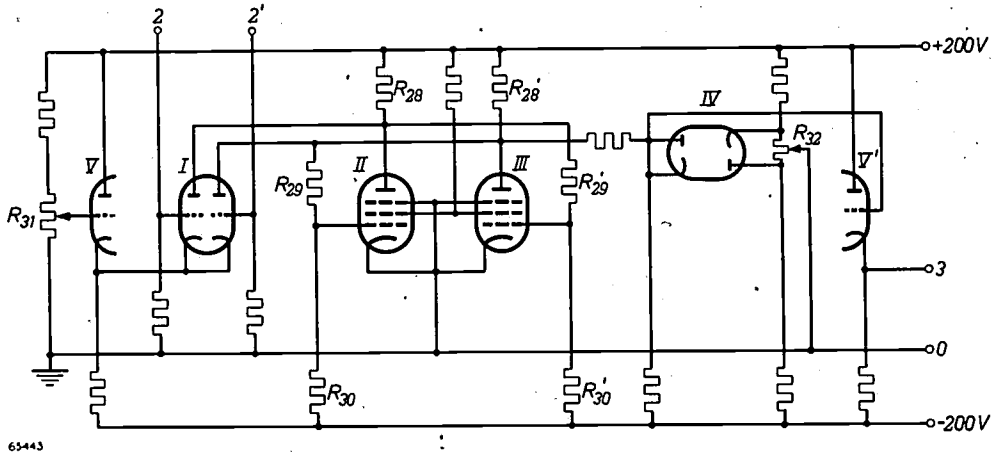


Fig. 22. Model of a controller with two-step action, in which the flip and flop circuit of fig. 21 is employed (valves *II* and *III* with the resistors $R_{28} \ldots R_{30}'$). The grid-voltage pulses triggering the circuit from one stable position to the other are derived from the valve *I*, the input terminals of which, *2*, *2'*, are connected to the output of the process analogue via the universal amplifier and the output collecting stage. With the potentiometer R_{31}, via the triode *V*, it is possible to vary the cathode potential of *I* and thus the backlash of the controller; the extent of the backlash can be read from a measuring element not drawn in this diagram. The voltage fluctuations at the anode of *III* are limited by the double diode *IV* and transmitted by the triode *V'* connected as a cathode-follower. The D.C. voltage level at the output terminal *3* is reduced to zero by means of the potentiometer R_{32}.

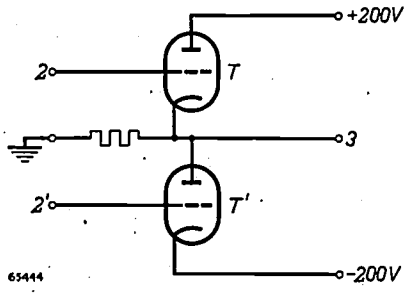


Fig. 23. Principle of a circuit corresponding to a controller with three-step action. The output terminal *3* is at a positive potential when the valve *T* is conducting, at earth potential when both valves are cut-off and at a negative potential when the valve *T'* is conducting.

When the output of this circuit is connected to an integrator one has an analogue of a controller with three speeds.

The cathode-ray oscilloscope

Vertical deflection

For the vertical deflection of the cathode-ray oscilloscope an amplifier is used which can be regarded as a greatly simplified form of the universal amplifier (fig. 13). The simplifications consist in: (1) the omission of the third stage (cathode-follower circuit), since in the present case the load of the amplifier is constant and has a high impedance, and (2) the omission of the voltage dividers between the two stages and between the second stage and the output, because in this case the D.C. voltage level at the output terminals need not be zero (it is here + 300 V, the same as that at the anode of the cathode-ray tube).

The gain is continuously adjustable by means of the resistor R_{24}, which has been made variable here (fig. 13).

Horizontal deflection

From a multivibrator synchronized with the mains voltage a voltage is obtained having the shape represented in *fig. 24* [9]). The amplitude of the step is 10 V and the average voltage level + 140 V. With the aid of a direct-voltage amplifier (one push-pull stage with two EF 42 pentodes) the amplitude is increased to about 100 V, with which an oscillogram can be traced across the whole width of the screen, and the direct-voltage level is raised to that at the anode of the cathode-ray tube (+ 300 V).

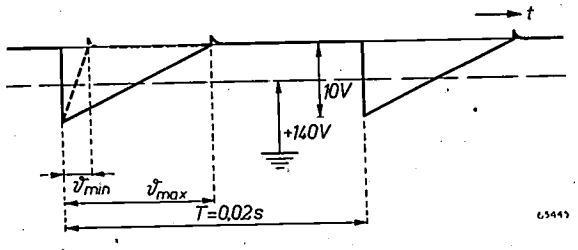


Fig. 24. Waveform of the voltage (taken from a multivibrator) which after amplification serves as time-base voltage of the oscilloscope. The duration of the stroke is variable between $\vartheta_{max} = 0.01$ sec and $\vartheta_{min} = 0.002$ sec.

[9]) Cf. J. van Slooten, Philips Techn. Rev. 12, 234, 1951 (No. 8), fig. 2.

The duration ϑ of the stroke is variable from 0.01 sec to 0.002 sec. A short stroke is chosen so as to spread a certain detail of the curve over the entire width of the screen. During the remaining part of the mains-voltage cycle the electron beam is suppressed by a square-wave voltage on the grid of the cathode-ray tube; this voltage is also supplied by the multivibrator.

When a short stroke is used the oscillogram is therefore traced only in a fraction of a cycle, and to give it sufficient luminosity for visual observation and photographic recording a cathode-ray tube with post-acceleration is used.

A small but practical detail is worth mentioning here. In article I it has already been shown how the step-response curve of the process analogue can be given a particular shape previously drawn on a sheet of cellophane, which is printed with a network of coordinates and placed in front of the oscilloscope screen. In the further use of the oscilloscope it is also desirable to have the network of coordinates visible together with the oscillogram. The background must then be much brighter than the darker lines of the coordinates, as is the case under ordinary circumstances but not when the light from the surroundings is cut off by a cylinder (K in fig. 12) with camera. Around the cathode-ray tube of the electro-analogue, behind the front plate, is a circle of small incandescent lamps uniformly illuminating the screen. Of course the brightness of the background must not be increased so far as to make the contrast with the luminescent trace too small. The current passing through the lamps is made variable so as to be able to adjust the brightness to the most favourable level.

Other parts

Fig. 25 shows a detail of the front of the electro-analogue. Here there are some parts which have not yet been discussed but which will now be dealt with briefly.

The square-wave generator has a fundamental frequency of 50 c/s. This square-wave voltage serves as input voltage for the process analogue when adjusting for a particular step-response, and also as a disturbance which can be introduced at various points in the control circuit. Part of the

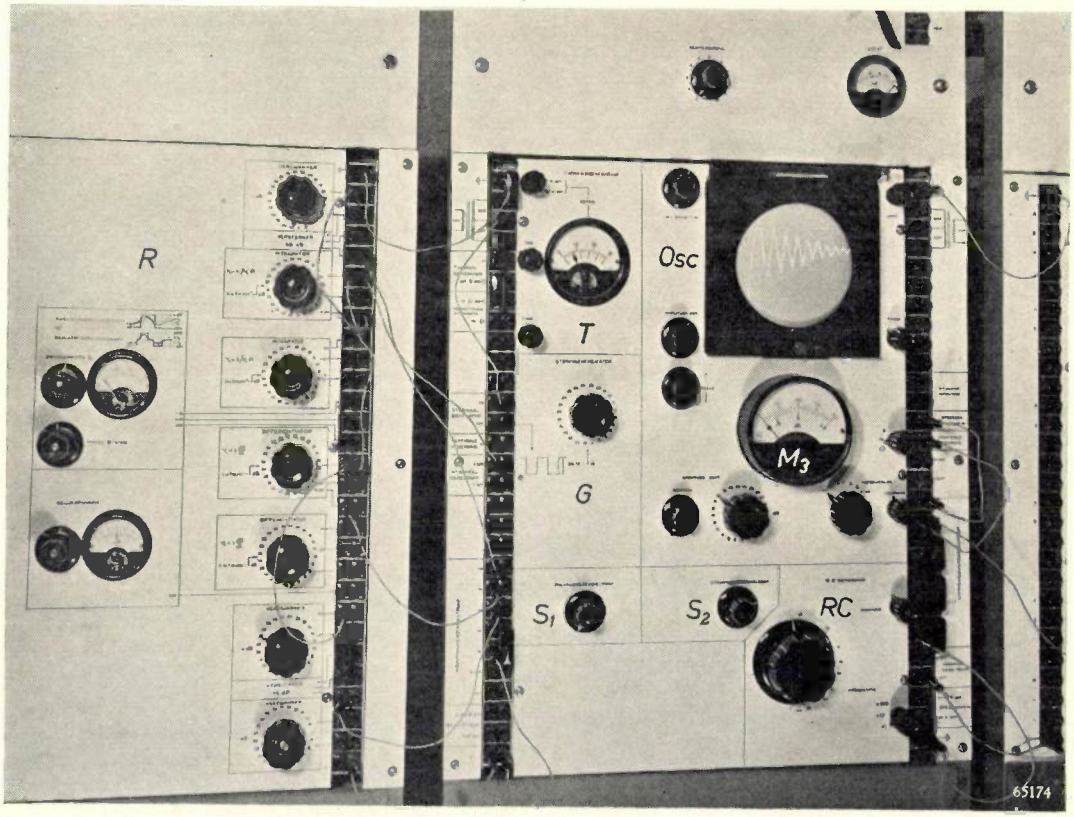


Fig. 25. Detail of the front of the electro-analogue. *Osc* oscilloscope with generator *T* of the time-base voltage, *G* square-wave generator, M_3 accuracy-measuring meter, S_1 input and S_2 output collecting stages, *RC* the *R-C* generator, *R* panel with controller models. Above the oscilloscope: control for adjusting the background brightness when photographing the oscillograms, and a micro-ammeter for measuring the intensity of the beam current.

circuit of the square-wave generator is shown in *fig. 26* and explained in the subscript.

For recording frequency-response curves and Nyquist diagrams, the electro-analogue is fitted with a sine wave generator. This is an *R-C* generator of a normal type with the frequency variable between 20 and 20,000 c/s.

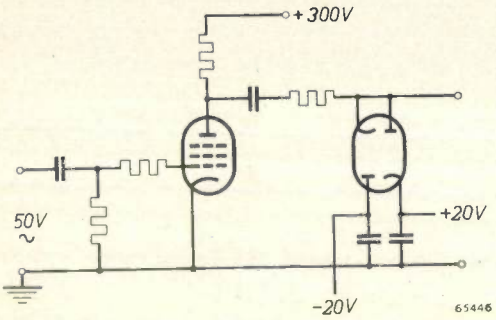


Fig. 26. A fairly large sinusoidal alternating voltage (50 V) alternately cuts off the pentode and makes it conducting, the anode voltage then varying between +300 V and about +10 V. This variation is limited by the double diode. In this way a square-wave voltage with an amplitude of about 15 V is obtained, which, attenuated or not, can be applied to the input of one of the collecting stages.

Finally mention is to be made of the performance meter with which one can determine how far the quantity to be controlled deviates from the value desired. As explained in article I, both the average of theabsolute values of x (x = the instantaneous value of the deviation) and the r.m.s. value of x can serve as a measure for this deviation. Both these quantities can be read from a moving-coil meter, the first after full-wave rectification of x by means of a bridge circuit of selenium cells, the second by the intermediary of a thermocouple.

The supply

All the direct voltages needed in the electro-analogue are derived from stabilised rectifiers, of which there is a fairly large number (37!), each supplying only one or several parts. In this way a source of instability is avoided which would arise if all parts were fed from one common source, and which would be difficult to eliminate. Moreover this solution has the advantage that stabilised rectifiers for small powers can easily be made in such a way as to yield a highly constant voltage while having a low output impedance; this greatly reduces the danger of all sorts of difficulties arising.

It was possible to manage with only a few types by choosing suitable direct voltages and currents and, where necessary, connecting two or more rectifiers in series. An idea of what the back of the electro-analogue looks like can be formed from *fig. 27*, showing the numerous supply rectifiers.

By way of an example the supply diagram of the oscilloscope is given in *fig. 28*.

Although for the working of stabilised rectifiers we may refer to an article already published in this journal [10]), in *fig. 29* the main features of the system have been indicated

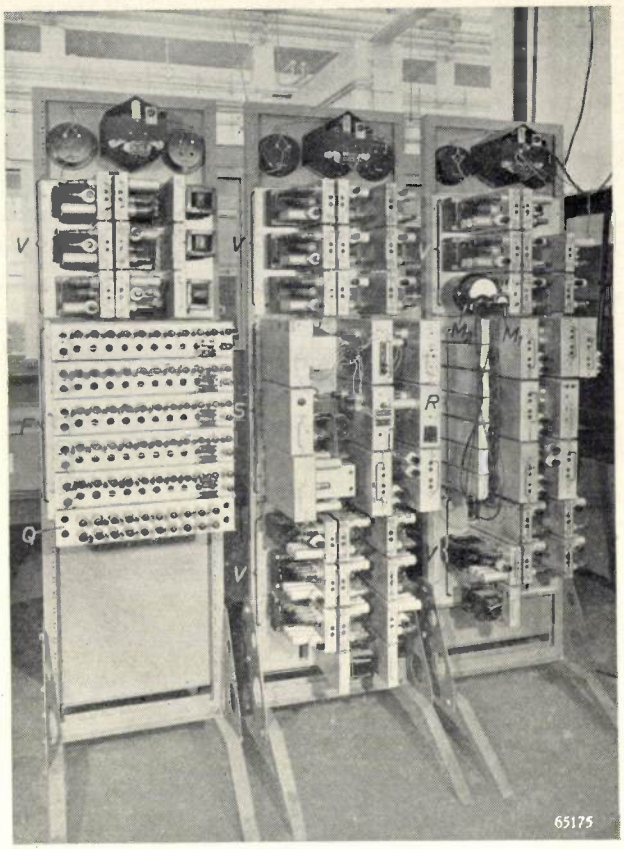


Fig. 27. Back of the electro-analogue. V stabilised supply rectifiers (37 in all), F delay network, Q special process analogues, S oscilloscope and various other parts, R parts of controller models, M_1 meter for checking the uniformity of the D.C. voltage levels (cf. subscript to fig. 13), M_2 D.C. voltmeter, which can be connected to the various rectifiers.

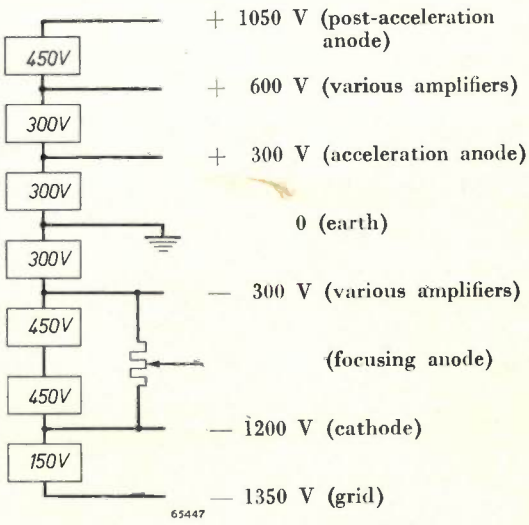


Fig. 28. Lay-out of the supply circuit for the oscilloscope.

[10]) H. J. Lindenhovius and H. Rinia, A direct current supply apparatus with stabilised voltage, Philips Techn. Rev. **6**, 54-61, 1941.

so as to be able to explain the essential elements of a regulating circuit [11]). The valves UL 41 connected in parallel form, as it were, a regulating unit the "position" of which is governed by the valve UF 41, which acts as controller and is influenced by the voltage that is to be kept constant. Here we have a proportionately acting controller. The offset is therefore not zero, but thanks to the high gain of the controller it is in any case very small.

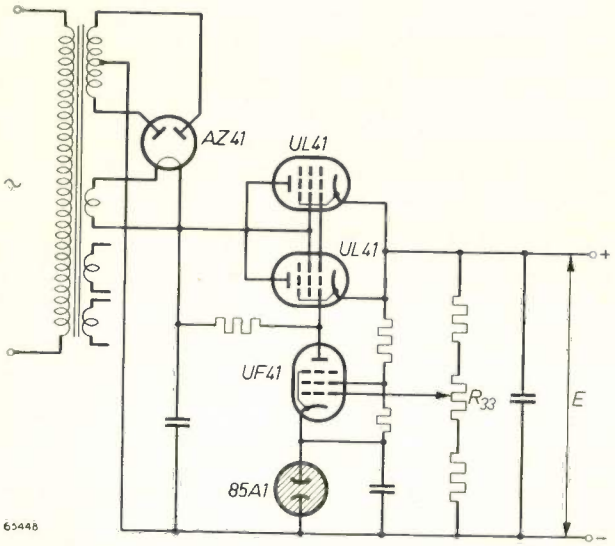


Fig. 29. Circuit diagram of a stabilised rectifier for 50 mA. AZ 41 rectifying valve, UL 41 parallel-connected pentodes used here as triodes ("regulating unit"), UF 41 amplifying valve ("controller"). The input voltage for the controller consists of the difference between a part (variable with the potentiometer R_{33}) of the direct voltage E and the stabilized voltage kept constant by the voltage reference tube 85 A1.

In addition to this type for 50 mA (and different voltages), also a type for 100 mA is used in the electro-analogue. This works with two parallel-connected AZ 41 valves and three parallel-connected UL 41 valves.

Mechanical construction

The components of the electro-analogue are mounted on interchangeable chassis, with the electrical contacts made by means of plug pins and sockets (*fig. 30*). For each type of part there is a spare duplicate which in a few moments can be put in to replace a defective part, so that in the case of a breakdown there is no need to wait until the cause has been traced and remedied. The construction of the chassis is such that all the elements mounted on them and all the wiring are easily accessible. With the chassis placed in the vertical position (see fig. 27) there is the least possible obstruction to the circulation of air.

Supporting the chassis are mounting plates screwed onto the back of the panels on which the electro-analogue is mounted. In front of these plates are the permanent leads forming the connections between the chassis and, among others, some rows

[11]) Cf. H. J. Roosdorp, On the regulation of industrial processes, Philips Techn. Rev. **12**, 221-227, 1951 (No. 8).

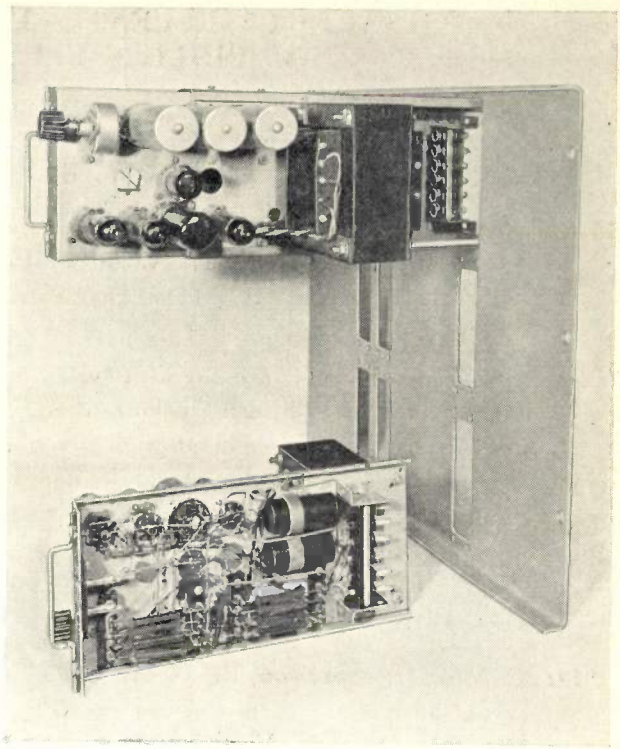


Fig. 30. Illustration of the interchangeability of the parts mounted on separate chassis. The mounting plate on the right can carry 9 chassis, one of which (at the top) with its guiding frame is shown here drawn out half way — this is the chassis of a supply rectifier for 50 mA. Below is an identical rectifier unit seen from the other side; note the easy accessibility of all elements, and the plug pins for making contact with sockets on the guiding frame. On the extreme left of the two chassis are sockets to which a voltmeter can be connected, and also the knob for adjusting the direct voltage to a certain value.

of plug sockets. By means of these plug sockets and separate leads the circuits required can be connected up. Finally the wiring compartments are shut off with front plates (fig. 25).

The construction described here has proved to be quite satisfactory for the assembling of a complicated system like that of the electro-analogue.

Summary. As a continuation to an article in which the parts used in making up the electro-analogue and the results that can be achieved with it have been described, the electric circuits of these parts are dealt with more closely. In the "planning" of the electro-analogue the fundamental idea was that it should be possible to get any desired combination of parts by plugging in separate connecting leads. All input and output terminals must therefore have the same direct-voltage level. The method usually followed for such a purpose — employing separating capacitors and leak resistors — cannot be considered in this case, because the large number of these *R-C* couplings which would have to be contained in the control circuit would lead to instability. Instead of these couplings, voltage dividers have been used, which provide for the correct direct-voltage level; not only positive but also negative supply voltages are needed for this, and moreover these have to be carefully stabilised.

Among the various parts discussed in this article are: the process analogues, the universal amplifier, the collecting stages, the integrator and the differentiator, models of discontinuously acting controllers, the oscilloscope, the supply and the mechanical construction.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1948*: W. de Groot and A. A. Kruithof: De kleurendriehoek (Ned. T. Natuurk. **16**, 235-241, 1950, No. 10). (The colour triangle; in Dutch.)

Description of a device consisting of a system of three slits provided with diffusing colour filters, a triangular screen and a triangle diaphragm covering a plate of opal glass. The colour triangle appears on the opal glass with the different colours in their proper places, when the three slits are illuminated by an incandescent lamp (See Philips Techn. Rev. **12**, 137-144, 1950, No. 5).

1949: J. Volger: Note on the Hall potential across an inhomogeneous conductor (Phys. Rev. **79**, 1013-1024, 1950, No. 6).

The change in the potential function in a homogeneous conductor of rectangular shape due to the Hall effect has been calculated to the first approximation. Some important conclusions are drawn for the case of an inhomogeneous conductor.

1950: R. van der Veen: Induction phenomena in photosynthesis, III (Physiologia Plantarum **3**, 247-257, 1950, No. 3).

It has been shown that the reaction of Chlorella to sudden illumination is much different from that of higher plants. While higher plants always begin with an initial uptake of CO_2, Chlorella does exactly the opposite and begins with a release of CO_2 before it starts photosynthesizing. This sudden release of CO_2 is positively correlated with light intensity, negatively with temperature and positively with the duration of the preceding dark period.

Another difference between higher plants and Chlorella is the "de-adaptation" in darkness. In Chlorella there is not much difference in adaptation time after short and after long dark periods. In higher plants, however, after long periods of darkness adaptation takes much longer than after short periods of darkness.

A third difference is the behaviour of Chlorella in pure hydrogen without any CO_2. In this environment higher plants react to illumination with a sudden releas of oxygen, which could not be found when examining Chlorella.

Thus Chlorella has a rather different pattern of photosynthesis than higher plants. Protococcus olivaceus Rabenh., which, like Chlorella, belongs to the Protococcales, behaves in the same way as Chlorella. Ulothrix sp, like Hormidium flaccidium and Stichococcus bacillaris, all three belonging to the Ulotrichales, have induction lines like those of higher plants.

1951: H. O. Huisman: Investigations on quinones and quinone derivatives. Preparation and antibiotic properties of some substituted p-benzo- and p-toluo-quinones, hydroquinones and hydroquinone esters (Rec. Trav. chim. Pays-Bas **69**, 1133-1156, 1950, Nos 9 and 10).

The antibiotic properties of the compounds named in the title have been tested in vitro against a yeast, various moulds and Gram-positive and Gram-negative micro-organisms. The toxicity has been determined in white mice by oral administration and by interperitoneal injection. The chemotherapeutic activity has been investigated in white mice injected with a lethal dose of pneumococci.

1952*: P. Cornelius: Electrische eenheden en het Giorgistelsel (Electrotechniek **28**, 454-460 473-479, 1950 Nos 23 and 24). (Electrical units and the Giorgi system; in Dutch.)

Review of the Giorgi system of units; see Nos **R 103, R 111, R 151** and Philips Techn. Rev. **10**, 55-60, 79-86, 1948.

VOL. 12 No. 12, pp. 337-364 JUNE 1951

Philips Technical Review

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

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

SPECTROCHEMICAL ANALYSIS

by N. W. H. ADDINK and W. de GROOT.

545.828:537.52

In an industry of any size there are always chemical analyses to be carried out, both for testing the raw materials required for manufacture and for various tests during the process of manufacture, as well as for purposes of scientific research. Apart from ordinary chemical analysis, spectroscopy is also applied, whereby the substance to be examined is placed in a flame, a spark or a carbon arc and caused to emit light. In this article a method of spectrochemical analysis is described in which the carbon arc is employed.

Introduction

When an element in gaseous form is heated to a high temperature or subjected to an electric discharge it radiates light consisting of a number of radiations of very specific frequencies characteristic for the particular element. When the light is examined with the aid of a spectroscope or spectrograph a number of sharply defined lines (spectral lines) are seen which together form the spectrum of the element.

The study of these spectra is of great importance in physics because it has given a deeper insight into the structure of the atom. Chemistry, too, has profited from this study, in that the spectrum constitutes a sensitive mears of identifying the kinds of atoms.

At first the spectroscopic method was used for a qualitative determination of the presence of a certain element, but at a later stage, as was understandable, attempts were made to use spectroscopy for quantitative analysis.

Kirchhoff and Bunsen, who, about 1860, first pointed out the importance of spectroscopy in chemistry, caused various elements to emit light by vaporizing a small amount of a compound, for instance a salt or an oxide, in a colourless gas flame (the Bunsen burner). Nowadays, for quantitative analysis, electric discharges in the form of a spark or an arc are frequently used, in addition to flames.

In this article a method of quantitative spectrochemical analysis will be described which employs a carbon arc fed with direct current. A weighed quantity (a few mg) of the sample to be tested is fully vaporized in the positive crater and the spectrum of the light emitted during this vaporization is photographed. From the blackening on the photographic plate caused by the various spectral lines it is possible to deduce the concentrations of the components of the sample. This method has been worked out to a routine method which is as accurate as and quicker than the ordinary chemical analysis.

Before proceeding to describe this method it is well to consider the ways in which a spectrum can arise.

The formation of a spectrum

The energy level diagram

The frequencies characteristic for an element can be plotted in what is known as an energy level diagram. As an example the energy level diagrams of the mercury atom and the neon atom are given in *fig. 1*, where the horizontal lines represent the energy levels for a number of stationary states (excited states) of the atom in question.

For the sake of clarity the higher energy levels have been omitted. These higher levels converge upon a certain limit, called the ionization energy, corresponding to the energy required to seperate the optical electron from the atom in its normal states.

The vertical lines denote "transitions" between the energy levels. According to Bohr (1913), in the transition from a state with energy E_1 to a state with energy E_2 ($<E_1$) a light quantum is emitted with the frequency ν, such that

$$h\nu = E_1 - E_2 . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where h represents the Planck constant. Inversely, an atom in the state with energy E_2 can pass, through absorption of a quantum of radiation of the same frequency ν, into the state E_1. The energies E_1 and E_2 are mostly measured from the ground level corresponding to the normal, non-excited state of the atom.

Putting $E = eU$, where e ($= 1.602 \times 10^{-19}$ coulomb) is the charge of the electron and U is the potential difference (in volts) that an electron originally at rest has to pass to get a (kinetic) energy E, then, considering the values of h ($= 6.625 \times 10^{-34}$ W sec^2) and c (velocity of light $= 2.998 \times 10^8$ m/sec), the relation between the potential difference $\Delta U = U_1 - U_2$ and the wavelength $\lambda = c/\nu$ is:

$$\lambda \cdot \Delta U = 12{,}395 \text{ Å} \cdot \text{V} .$$

In accordance with the interpretation of the energy level diagram, to cause an atom to radiate light with a certain frequency it has to be excited with an energy equal to the value E_1 of the starting level. Except that E_2 together with E_1 determines the frequency — according to eq. (1) — the value of E_2, the energy of the final level, does not play any part here. The truth of this statement has been conclusively confirmed by the experiments carried out by Franck and Hertz [1]).

Fig. 2 gives an example of this. Electrons having a constant velocity v obtained by causing them to traverse a potential difference U, so that

$$\tfrac{1}{2} m v^2 = e U ,$$

are shot into a space containing neon, under a pressure of 0.5 mm, and a little mercury vapour. An atom can be brought to a state with energy E_1 if

$$eU \geqq E_1 . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In accordance with the diagram in fig. 1, with increasing U there appear first the mercury lines in

[1]) See, e.g., Handb. d. Physik (Geiger und Scheel) Bd. 3, I, 23—184, Springer, Berlin, 1933.

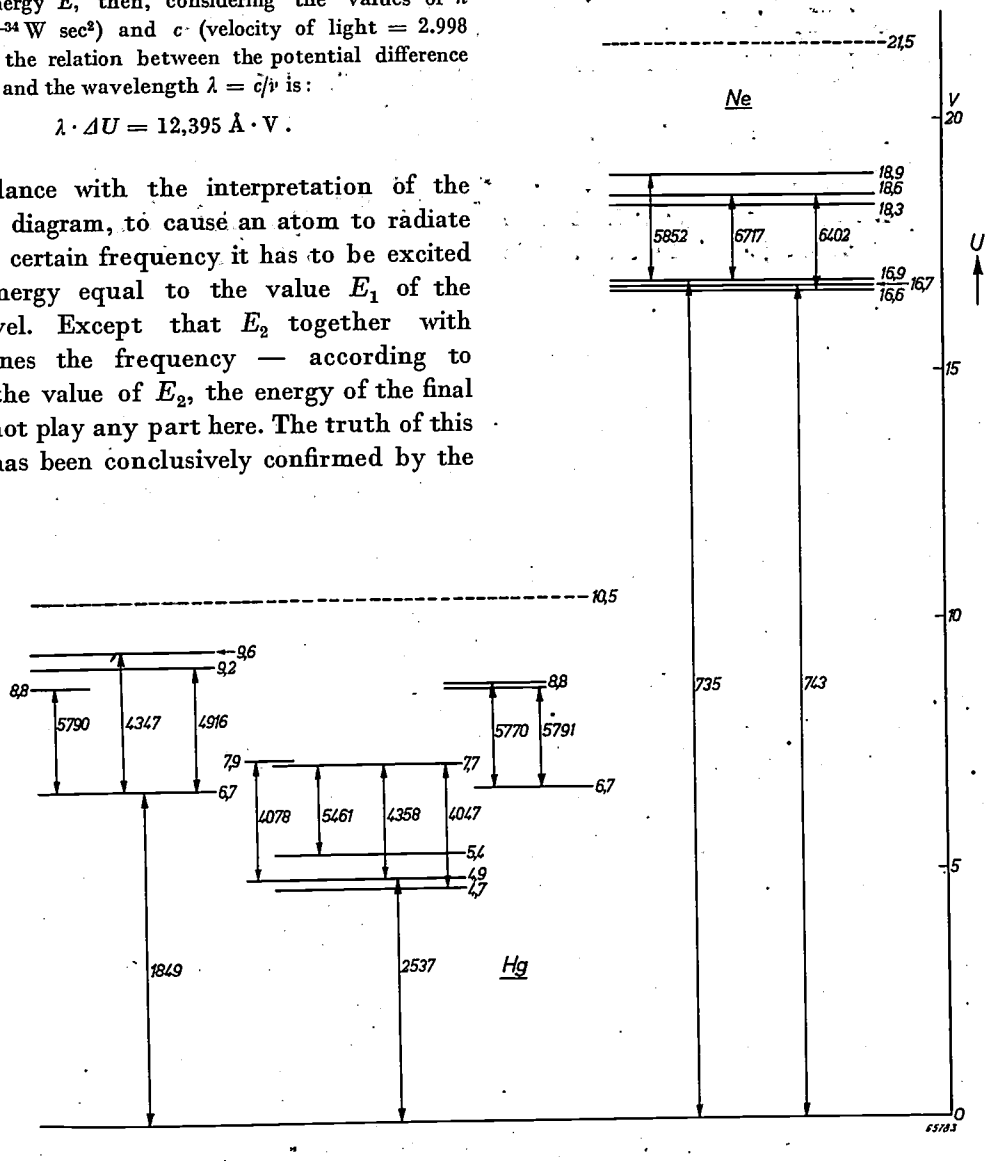


Fig. 1. Energy level diagram (incomplete) of mercury (Hg) and neon (Ne) with some transitions. The wavelengths corresponding to the transitions are given in Å.

order of their initial levels and then the red neon lines.

This method of obtaining a spectrum with the aid of electrons with constant velocity is of great value when investigating the "structure" of a spectrum. For chemical analysis, however, it is less suitable.

rium is brought about by a number of pairs of antagonistic processes [2]), viz :

1) absorption and emission of light;
2) formation and destruction of excited atoms through collision with electrons;
3) ditto through collision with other atoms or molecules that may be present.

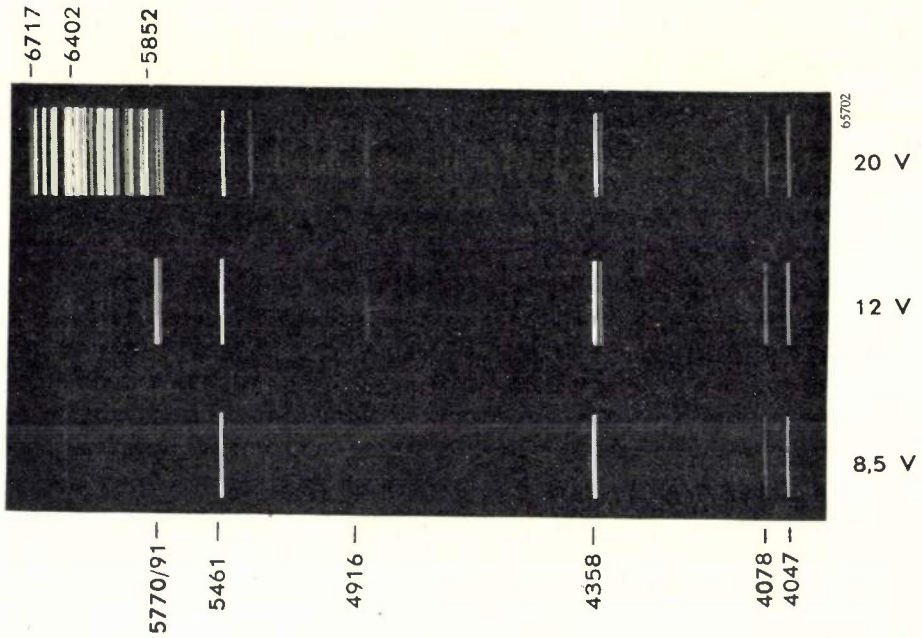


Fig. 2. Spectrum of a mixture of mercury and neon bombarded with electrons accelerated by voltages of 8.5 V, 12 V and 20 V. The weak lines on the left in the spectra for 12 V and 8.5 V are neon lines produced by scattered light from a part of the test tube in which fast electrons are present.

Furnace spectra

A spectrum can also be obtained by introducing an element in the gaseous or vapour form into an entirely closed space the walls of which have been brought to a constant temperature T. Thermodynamics teach us that in this case conditions are entirely determined by the value of T and that the number of atoms of a certain kind, in a state with energy E, is given by

$$\frac{n_E}{n_0} = \frac{g_E}{g_0} e^{-E/kT} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where n_E is the number of atoms in the excited state and n_0 the number of atoms in the ground state; the g's are the weight factors characterictic for the element and for the respective states, which factors are small integer numbers.

From the kinetic point of view this state of equilibrium

Suppose a small opening is made in two opposite walls of the enclosed space so that one can see through it, or, as King [3]) did, a furnace is used (see *fig. 3*) in the shape of a tube open at both ends and with its wall brought to a constant high temperature. Then, upon a substance, say iron, being vaporized in that space or tube one can see the spectrum of that material appear. The intensity of each spectral line is proportional to the number of excited atoms n_E. Upon the temperature being raised the number of atoms with higher energy increases. At low temperature only the lines starting from levels with low energy are to be seen. An example of this is given in *fig. 4*.

[2]) It is to be stressed that, according to the principle of detailed equilibrium, each pair of antagonistic processes is in itself capable of bringing about the equilibrium. See, e.g., M. J. Druyvesteyn and W. de Groot, Physica **12**, 153—166, 1932.

[3]) A. S. King, Astrophys. J. 27, 353, 1908.

In order to predict how different elements will behave in this respect, a table of ionization voltages can be used; the ionization energy $E_i = eU_i$, as upper limit, is a rough measure for the excitation

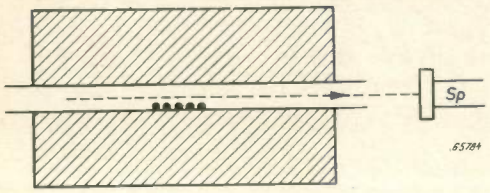


Fig. 3. King's furnace. Sp= spectograph.

energy. Such a table is given below (*table I*) and graphically represented in *fig. 5*, from which it is seen that the ionization voltage U_i of most elements lies between 6 and 10 volts. Ionization potentials less than 6 volts occur in the case of alkali metals and further for a few elements such as Sr, Ba, La, Hf, Ra and U, thus mainly in the left-hand bottom corner of the table. Ionization potentials higher than 10 volts are found in the case of the rare gases, the halogens and the elements O, S, N, P, As, C and Hg, thus mainly in the right-hand top corner of the table.

The same processes as mentioned above will take place likewise in the flame; the pairs of processes (2) and (3) will be fully active, but not so the pair (1), because though there is radiation outward there is no radiation inward. It can be proved, however, that the latter has little effect upon the equilibrium as a whole, so that the intensities depend upon the temperature in much the same way as in the case of the furnace. The temperature of the flame from an ordinary Bunsen burner is not particularly high, about 1700 °C, so that it is suitable only for elements which are readily evaporated and have a relatively low ionization voltage. In spectrochemistry special flames are now often used which are fed with a mixture of gas and oxygen and with which higher temperatures are reached.

Electric discharges at low pressure

Apart from exposing a gas or vapour to a high temperature, it can be caused to emit light by means of an electric discharge, as intimated when referring to the experiments of Franck and Hertz. Here a distinction has to be made between discharges under low pressure and with low current density

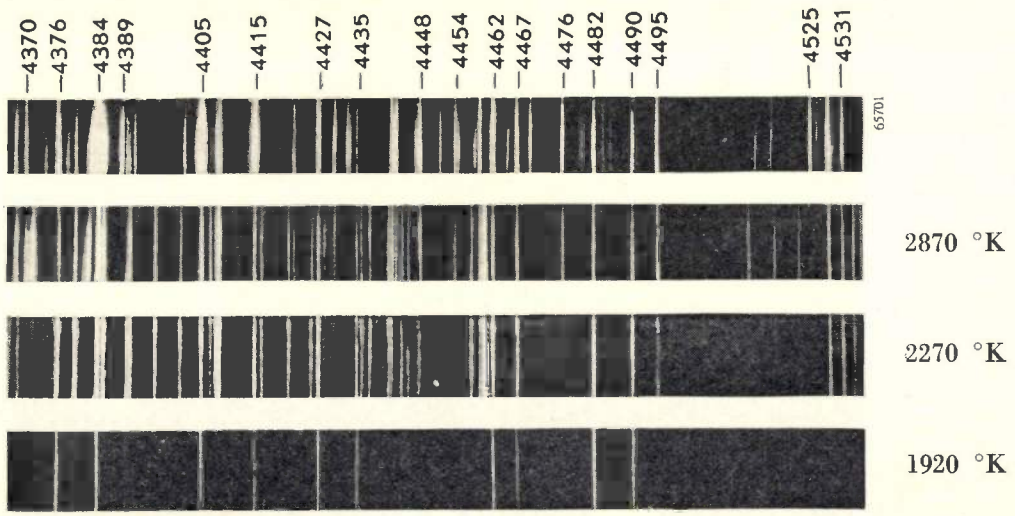


Fig. 4. Photographic record, according to King, of the spectrum of iron at three temperatures. Above these, the arc spectrum of iron is shown.

With the exception of the elements mentioned, therefore, most kinds of atoms will be visible at about the same temperature. A temperature of 2000 to 3000 °K is sufficient.

The flame

Instead of heating them in vapour form in a furnace, the elements can be exposed to a high temperature in a flame which in itself is not luminous.

(Geissler tube) and those under high pressure and with high current density (arc and spark).

Discharges in gases of low pressure have played an important part, for instance, in the discovery of the rare gases. The use of gas discharges of low pressure for generating spectra [4]) is almost as old

[4]) See, e.g., W. Elenbaas and J. Riemens, Light sources for line spectra, Philips Techn. Rev. **11**, 299—302, 1950 (No. 10).

as that of the Bunsen flame. Mostly the discharges have a positive column, in which the electrons are accelerated, but no farther than to the velocity at which they have gathered sufficient energy to excite and ionize atoms. Thus there is a limit to which the electron energy can be raised. As a result the velocity distribution appears to correspond to Maxwell's distribution with a fictitious temperature T_e (electron temperature). This temperature is of the order of 1500 U_i and thus,

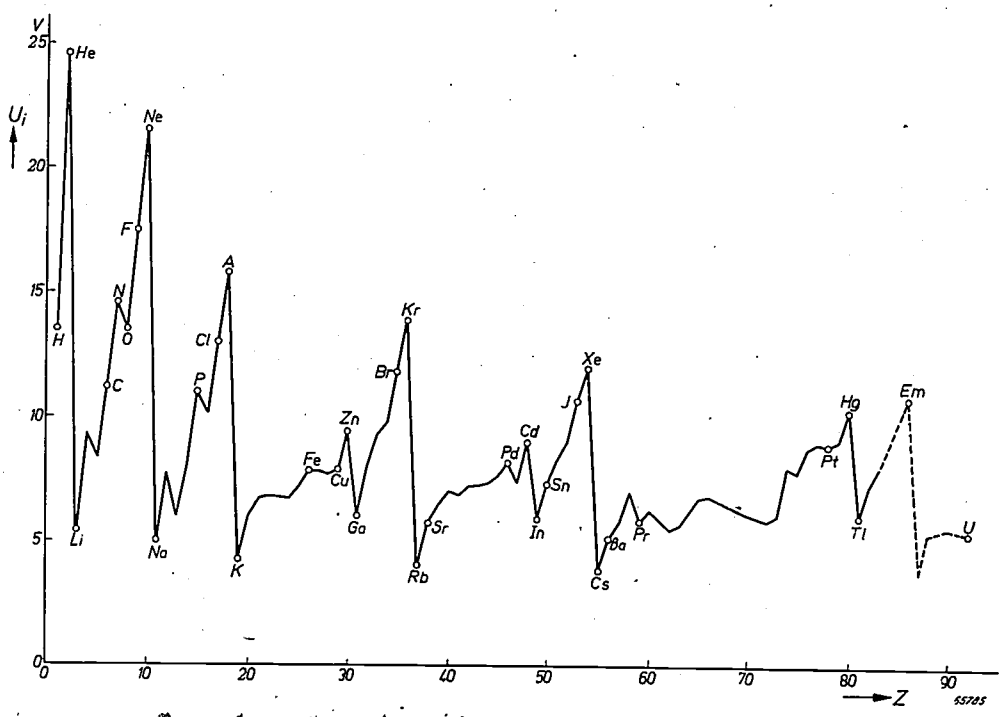


Fig. 5. The ionization potential U_i of the elements, plotted as a function of their atomic number Z.

Tabel I. The chemical elements arranged according to the periodic system. Above each element is the atomic number and underneath the ionization potential in volts.

1 H 13.6																2 He 24.6
3 Li 5.4	4 Be 9.3	5 B 8.3	6 C 11.3										7 N 14.5	8 O 13.6	9 F 17.4	10 Ne 21.6
11 Na 5.1	12 Mg 7.6	13 Al 6.0	14 Si 8.1										15 P 11.0	16 S 10.4	17 Cl 13.0	18 A 15.8
19 K 4.3	20 Ca 6.1	21 Sc 6.7	22 Ti 6.8	23 V 6.7	24 Cr 6.8	25 Mn 7.4	26 Fe 7.9	27 Co 7.9	28 Ni 7.6	29 Cu 7.7	30 Zn 9.4	31 Ga 6.0	32 Ge 8.1	33 As 10	34 Se 9.8	35 Br 11.8	36 Kr 14.0
37 Rb 4.2	38 Sr 5.7	39 Y 6.6	40 Zr 6.95	41 Nb 6.8	42 Mo 7.4	43 Tc	44 Ru 7.5	45 Rh 7.7	46 Pd 8.3	47 Ag 7.6	48 Cd 9.0	49 In 5.8	50 Sn 7.3	51 Sb 8.6	52 Te 9.0	53 J 10.4	54 X 12.1
55 Cs 3.9	56 Ba 5.2	57 La 5.6	72 Hf 5.5	73 Ta 6	74 W 8.0	75 Re 7.9	76 Os 8.7	77 Ir 9.2	78 Pt 9.0	79 Au 9.2	80 Hg 10.4	81 Tl 6.1	82 Pb 7.4	83 Bi 8	84 Po	85 At	86 Em 10.7
			58 Ce (6.9)	59 Pr (5.8)	60 Nd (6.3)	61	62 Sa (5.6)	63 Eu 5.7	64 Gd 6.2	65 Tb (6.7)	66 Dy (6.8)	67 Ho	68 Er	69 Tm	70 Yb 6.2	71 Cp 5.0	
87 Fa	88 Ra 5.3	89 Ac	90 Th	91 Pa	92 U 4	93 Np	94 Pu	95 Am	96 Cm							

for instance, in the case of neon (U_i = 21.5 V) about 30,000 °K. On the other hand the gas itself is relatively cold (e.g. <100 °C). Consequently in this discharge only the pairs of processes (1) and (2) are active and owing to the absence of inward radiation there is again no complete equilibrium of radiation. Therefore, as a general rule, the relative intensities of the various spectral lines cannot be deduced from a temperature equilibrium. In the case of a gas mixture the spectrum of a substance with high ionization potential can be entirely suppressed by a substance with a lower ionization potential, owing to the fact that the electron temperature is determined by the component with the low ionization potential, provided, of course, that this component is present in sufficient concentration. A typical example is the sodium lamp which contains neon gas for starting the discharge in the cold state. Upon ignition this lamp gives a red neon light and as the temperature of the lamp rises this light is gradually replaced for the greater part by yellow sodium light, while the electron temperature drops from the high level it assumes in pure neon to a much lower level (about 9000 °K) determined by the sodium.

In the quantitative analysis of gas mixtures with the aid of low-pressure discharges the foregoing has has to be taken into account, so that it can only be used, for instance, for comparing mixtures of about the same composition.

Discharges at high pressure; arc and spark

The arc discharge (discharge of great current density, with thermionic emission of the cathode) in gases of atmospheric pressure is distinguished from the form of discharge discussed above in that the gas assumes a high temperature. This is due to the fact that the elastic collisions (collisions not accompanied by ionization of excitation) between electrons and gas atoms or gas molecules are so numerous that the gas receives a considerable quantity of kinetic energy which can be carried off only by means of thermal conduction and radiation. It may be taken that also in the arc the electrons have a Maxwellian velocity distribution. The electron temperature, however, will be only slightly higher than that of the gas. The condition in the arc is therefore comparable to that in the flame, with this difference that the temperature is usually much higher. Consequently the arc, and particularly the arc between carbon electrodes, is exceptionally suitable as a source of light for spectrochemical analysis: owing to the high temperature of the electrodes almost all substances can be vaporized in the arc, and with the high temperature of the gas also spectral lines of elements with a fairly high ionization potential are made visible.

The suppression of the spectrum of a not readily ionizable element by that of a readily ionizable element also occurs here to a certain extent, because an element with low ionization potential is strongly excited and thus gives a strong radiation, which causes the arc temperature to drop. For quantitative examination where it is desired to produce all the spectra, as far as possible under equal conditions, it is therefore necessary that the substance to be examined should be present in only a small concentration, so that it will not greatly affect the temperature of the arc.

In addition to the arc discharge, also the spark is much used in spectrochemistry. A spark is an instantaneous discharge in a narrow conducting channel formed in the gas between the two electrodes. The conditions in the sparking channel are comparable to those in the arc discharge but the electrodes are for the greater part cold, only the cathode being locally heated to a high temperature. There is extensive literature available about the use of the spark for quantitative analysis [5]). We shall not go into this here because in this article we wish to confine ourselves to the method where the carbon arc is employed, fed with direct current.

The reasons why the arc is preferred to the spark lie in the following points:

1) Since in the spark method of analysis only very small quantities of the substance are vaporized (0.05 to 0.5 mg), whereas with the arc method 5 or 10 mg is vaporized and brought to emission, the latter method is to be preferred in cases where the material to be examined is not homogeneous (e.g. magnet steel, silicon iron and suchlike).
2) The spark can only be used for special routine analysis with materials having the same antecedents (the degree of hardness of steel, for instance, affects the results of the analysis obtained with the spark as regards the secondary components in this material).
3) The method of arc analysis is more universal since it permits of metals, oxides and compounds in general being analysed in the same way, whereas for analysing non-conducting materials with the spark certain artifices have to be employed.

The carbon arc in air is particularly suitable for this analysis by reason of the fact that the main mass formed by the vapour cloud may be considered to be of a constant composition (O, N and C).

Relation between intensity and concentration

We now have to discuss the important question in how far the intensity of a spectral line observed,

[5]) See, e.g., W. Gerlach et al, Die chemische Emissions-Spektralanalyse I, II and III, Voss, Leipzig 1936.

as expressed in the blackening of the photographic plate on which the spectrum is photographed, can be taken as a measure for the amount of the respective element in the sample under test. This is a question which resolves itself into two others:

1) Is the concentration of the element in the arc proportional to the amount of that element in the sample?
2) Is the measured intensity of the line proportional to the concentration of the element in the light-emitting gas?

The concentration of the element in the arc is determined by the rate at which the element evaporates out of the sample. Thus it will depend not only upon the amount of the element but also upon other factors, such as the temperature-dependency of the vapour pressure of the sample. Given the same quantity, the concentration in the arc will therefore vary as between one case and another. This difficulty is overcome by arranging for a small quantity of the sample to be fully vaporized in the positive crater. If the duration of the photographic exposure is extended over the time taken for the whole of the sample to evaporate, then every atom in the sample has an equal opportunity to emit light and the intensity integrated by the photographic plate is independent of the aforementioned individual differences in the rate of vaporization. Moreover, if so desired, the vaporization can be retarded (by adding a material that does not readily evaporate, such as quartz) or accelerated (by adding a material comparatively easy to evaporate, such as nickel). Thus the line intensity observed is entirely determined by the (constant) temperature of the arc and by the excitation voltage.

With this method where the sample is brought to complete vaporization the material to be evaporated is in a cavity of the lower electrode functioning as anode, with the cathode placed vertically above it. Under the influence of convection in the gas the atoms move vertically upward, after which they ultimately leave the field of the discharge. This vertical motion is accentuated by some of the atoms being present in the form of positive ions and thus being drawn by the electric field of the arc towards the cathode, where they give off their charge. This results in an accumulation of atoms in front of the cathode, so that in its vicinity, in the so-called cathode layer, the emission of light is stronger than in the rest of the arc. For this reason the light from the cathode layer is not usually used for the quantitative spectral analysis; one places in front of the arc a screen with a horizontal slit through which only the light from the middle of the arc discharge can pass (see *fig. 6*). The cathode light is used only in certain cases, for instance when examining elements which are present in very small concentrations, in which case the lower electrode containing the material to be vaporized is often made to function as a cathode and the slit is omitted.

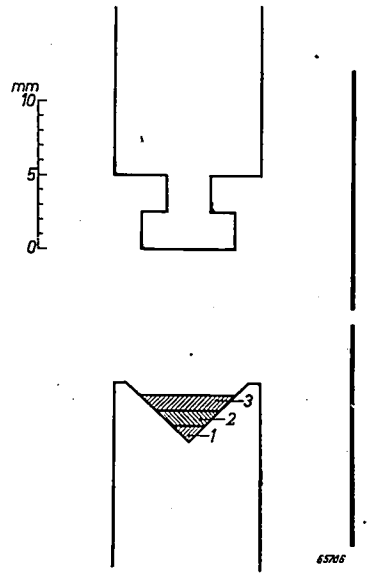


Fig. 6. Arrangement of the electrodes. The lower one, containing the specimen to be vaporized, is usually the anode. On the right the slit though which the light from the middle of the arc is allowed to pass to the spectrograph. In the case of a substance that is not readily vaporized the charge consists of *1* nickel. *2* the specimen, *3* graphite; in the case of a fairly volatile substance the charge is *1* specimen, *2* quartz, *3* graphite. The electrodes have been given their particular shapes to ensure a symmetrical form of the discharge.

We now come to the second question, viz. whether the line intensity is proportional to the concentrations of the atom contained in the gas. Under the conditions just described the same fraction of the element in a certain state of excitation will have been present in the part of the arc used for the analysis, so that one may expect the total amount of energy radiated in a certain spectral line to be proportional to the total number of atoms of that element present in the sample.

There is, however, one disturbing factor, the self-absorption. When a spectral line is formed by the transition from a state *1* to a state *2* then a quantum of emitted light meeting another atom in the state *2* may bring the latter into the state *1*; this atom is then itself capable of emitting a quantum of the same frequency. In this phenomenon, which is also called resonance, there is, therefore, no loss of energy in itself. The quantum of light, however, is retained longer in the discharge and the radiation is reduced, since only the atoms in the outermost layers of the gas can emit their quanta

of light unchecked and those in the excited state inside the arc are apt to be destroyed by one of the collision processes (2) or (3).

By a somewhat schematic argumentation this can be explained, for a transition to the ground level, as follows:

Let n_0 be the number of atoms in the ground state and n_E the number of excited atoms in a state E. The number of excited atoms formed by collision can then be represented by Cn_0 and the number destroyed by collision by $C'n_E$, whilst the number of emitted light quanta is represented by An_E and the number of excitated atoms reformed through self-absorption by Zn_0, all with reference to the unit of time. In the stationary state

$$n_0 (C + Z) = n_E (A + C').$$

The number of quanta leaving the discharge space per unit of time is:

$$Q = n_E A - n_0 Z \approx n_0 \left(\frac{CA}{C'} - Z\right)$$

since $n_E/n_0 \approx C'/C \approx \exp(-E/kT)$.

Since Z is proportional to the density of radiation, which very roughly is proportional to n_0, it follows that to a first approximation Q is a quadratic function of n_0:

$$Q \approx an_0 - bn_0^2.$$

The resonance absorption, which always occurs whenever the state *2* coincides with the initial state, increases with the concentration of atoms in the state *2*. Therefore, where the concentration is large, the line intensity will increase less than proportionately with the concentration of the excited atoms and thus the integrated line intensity will no longer be proportional to the total number of vaporized atoms.

In the case of transitions where the final level itself is a high-energy level there is usually little self-absorption. Preferably, therefore, lines situated "high" up in the energy level diagram will be used for the investigation. But this is not always possible. Where it is a matter of determining small quantities of an admixture one is usually restricted to principal lines, which are mostly transitions to the ground level and thus show self-absorption. Below it will be shown how in such a case a correction can be made for the effect of self-absorption.

How the tests are carried out

Fig. 7 shows the apparatus employed for the analyses carried out in the Philips Laboratory at Eindhoven.

As already stated, a weighed amount of the sample (5 mg), if necessary with the addition of an admixture for controlling the evaporation, is placed in the positive crater. The arc is then struck and the spectrum photographed. The distance from the arc to the slit in the spectrograph and the width of the slit (say 0.025 mm) are not varied. Of course the same photographic material is used and the photographic plates are all developed at the same temperature. Fluctuations in the photographic material, which cannot always be avoided even when it is of a perfectly reliable make, are regularly checked by taking special photographs.

After the photographic plate has been developed, fixed and dried the further work is done in two stages:

a) the qualitative examination, and

b) the measurement of the intensity.

In the qualitative examination it has to be determined what elements are present in the sample. In many cases this is known in advance, but then the lines of the various kinds of atoms on the plate have to be identified and those most suitable for measuring have to be picked out. Mostly the lines are identified by ascertaining how the lines are situated with respect to a fixed reference spectrum which is always photographed at the same time on another part of the plate. Mostly one chooses for this the spectrum of the iron arc, the wavelengths of which are accurately known.

Usually the spectrum is qualitatively examined by placing the plate in an apparatus which projects onto a white screen an enlarged image of part of the spectrum together with the reference spectrum (*fig. 8*).

Let us suppose that a certain line has been chosen for measuring its intensity. The classical method is that whereby the density is measured. The plate is placed in an apparatus by means of which a fine line of light is projected on the photographic picture of the spectral line. With the aid of a photocell and a galvanometer one then determines the quotient of the intensity I that has passed through the photograph of the spectral line and the intensity I_0 that has passed through a part of the plate where there is no image. The quantity $D = \log I_0/I$ is called the density. From this density the intensity has to be determined of the light that caused the blackening. This is done, as is known, with the aid of a number of density marks obtained by exposing the plate to a constant light source, both unattenuated and attenuated in certain ratios (with the aid of a rapidly rotating sector). The density of these marks is plotted against the logarithm of the intensity, thus giving a curve which is characteristic for the kind of plate and the method of developing.

When measuring intensities in this way one has to reckon with the blackening of the background, since in spite of all precautions some scattered light always falls upon the plate and slightly fogs it. This fogging

merges with the blackening caused by the line and allowance has to be made for it by separately measuring the background density, preferably close to the line to be measured, and then correcting the line intensity measured for the background intensity. Thus a great deal of labour is involved in this method of measuring intensity and as a consequence it takes rather a long time.

tative examination. The density marks on this s.p.d. scale are a number of lines which in width and sharpness resemble the spectral lines as projected onto the screen, these lines being obtained by a photographic process whereby their reflection density D^r (the logarithm of the ratio of the reflection of the white paper and of the density mark) is made to increase in equal steps (about 0.025).

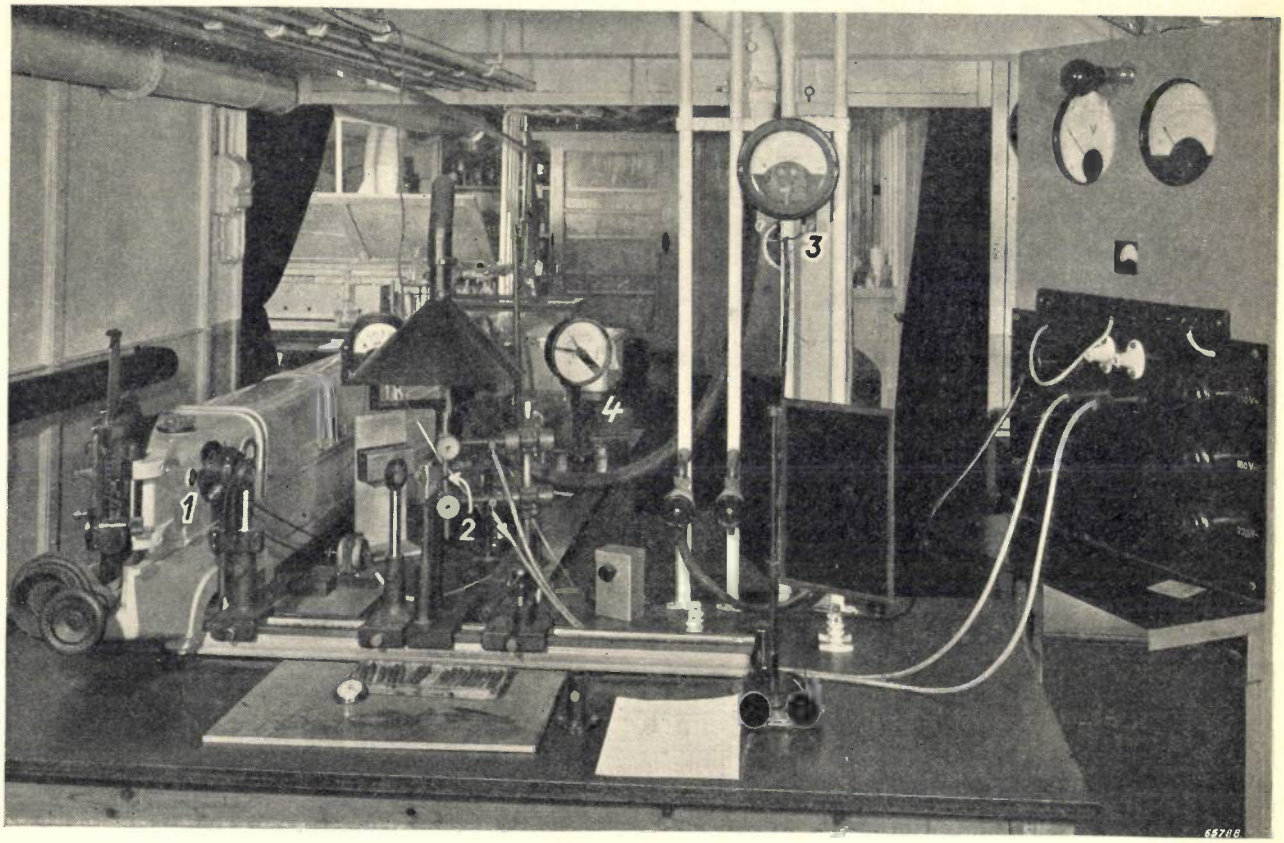


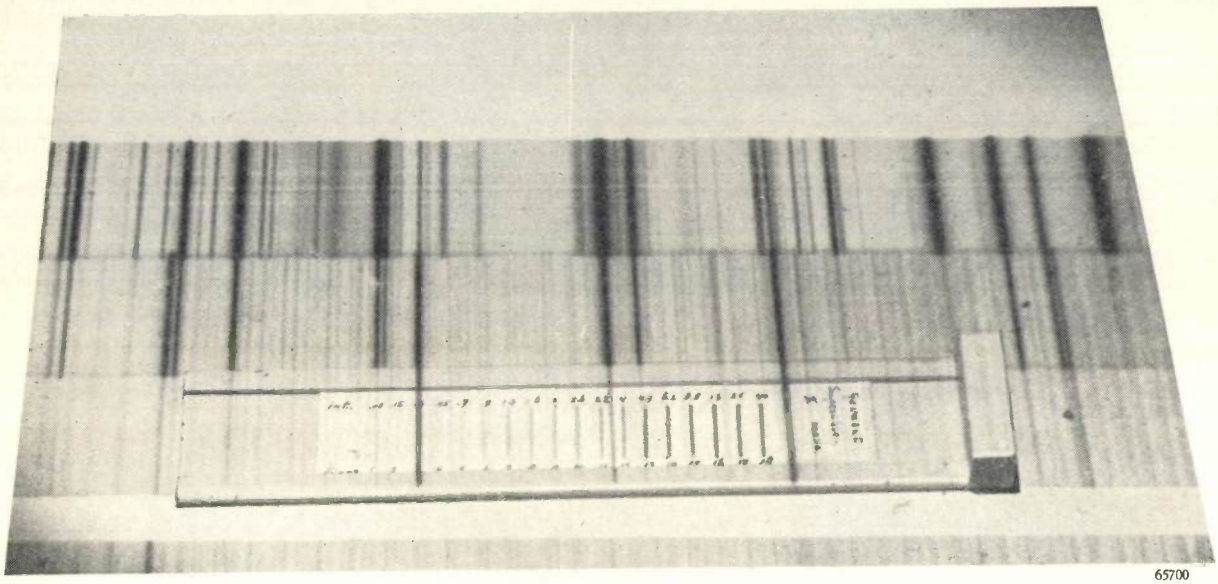
Fig. 7. Photograph of the whole of the apparatus employed for recording the spectra. On the left the automatic quartz spectograph (Hilger E492) with cassette open. On the optical bench at *1* is the spectograph slit, in front of which is a rotating sector and a focusing lens casting the light on the prism; at *2* the carbon arc with the horizontal slit, above which is a suction device. To the left of the slit is an auxiliary lens projecting an image of the arc on a screen beside the spectrograph slit, by means of which the source of light can easily be aligned in the optical axis. Behind this lens is the thermostat keeping the temperature of the room constant at 21-22 °C. At *3* the ammeter for reading the arc current (10 A). At *4* a torsion balance for quickly weighing off 5 mg of material. In the background apparatus for carrying out simple chemical manipulations.

A simplified method has therefore been introduced which saves time, inter alia, owing to the correction for the background density taking place automatically [6]).

What is known as an s.p.d. (standard paper density) scale — a strip of white paper bearing a number of density marks produced on it photographically — is laid on the horizontal white screen (fig. 8) onto which the spectrum is projected for the qualiThe weakest line has a density of about 0.075 (84% reflection) and the darkest about 0.375 (52% reflection). These lines are numbered from 1 to 13. The scale is shifted along the screen until the projected image of the spectral line to be measured is flanked by two successive lines of the scale, such that the density to be measured lies between the flanking densities. The density of the projected image can then be expressed in a number of the scale, and after some practice divisions of a step can be estimated.

[6]) See also N.W.H.Addink, Spectrochimica acta 4, 36—42, 1950 (No. 1).

Not only does this combine the qualitative with the quantitative test into one process, but at the same time it gives automatically a correction for the background density. This can be explained in the following way: Let the density of the projected line image on the white paper of the scale (logarithm of the ratio of the brightness of the white paper illuminated



65700

Fig. 8. The screen on which a magnified image of the spectrum is projected and upon which the s.p.d. scale is laid.

respectively through the non-exposed and through the exposed plate) be D_{l+b}, where l refers to the line and b to the background. The density of the line appears to correspond to a certain mark s on the scale. This mark, however, is illuminated by light which has passed through the fogged part of the negative beside the line, so that its density is $D^r{}_s + D_b$ (by the logarithmic definition the densities are additive). When the background density is not too intensive it has been found from experience that D_{l+b} may be replaced by $D_l + D_b$. By equation we get:

$$D_l + D_b \approx D_{l+b} = D^r{}_s + D_b ,$$

or

$$D_l \approx D^r{}_s .$$

Thus we get by direct means approximately the density, D_l, of the line without background density! All that is needed is to establish the relation between the number of the s.p.d. scale and the intensity of the spectral lines to be measured, by means of a scale of density marks on a photographic plate, and this need be repeated only now and again to check the constancy of the plate material.

Once the line intensity has been determined a relation has to be found between that intensity and the amount of the element concerned contained in the sample. This is done empirically by starting with specimens of known composition.

This does not imply that the spectrochemical analysis is entirely dependent upon the chemical analysis. In fact the concentration can be determined entirely spectrochemically by the method of addition. One adds to the sample containing an unknown quantity ($c=x$) of an element, different known quantities ($c-x = m, n, \ldots$) of the same element, and the intensities derived from the densities found are then plotted as functions of the quantities added. When the line connecting these points (see *fig.9)* is extrapolated to zero intensity the quantity sought is found directly along the (negative) abscissa [7]).

Different spectral lines will be chosen according to the quantity under examination, since it is preferred to have the reflection density on the screen lying in the interval 0.1—0.3 (50—80% reflection), because there the contrast sensitivity of the eye is greatest.

Where just one single measurement is to be taken the work entailed in the calibration of the scale takes more time than the actual measuring, but this is of no consequence in the case of routine

[7]) See N. W. H. Addink, Rec. trav. chim. Pays-Bas **67**, 690—696, 1948; **70**, 155-167, 1951 (No. 2).

examinations where a large number of analogous cases have to be investigated and the same calibrated scale can be used for them all.

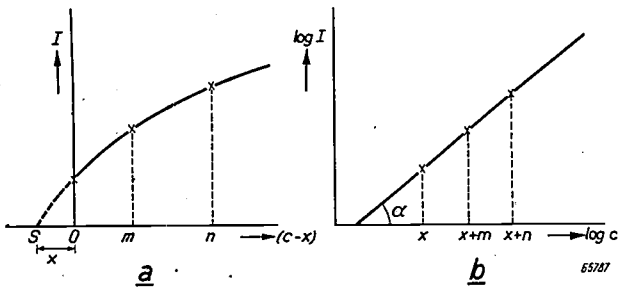


Fig. 9. Explanation of the additive method for determining concentrations.
a) By plotting the intensity I as a function of the amount added $(c-x)$, after extrapolation the point S is found. The distance SO is the unknown concentration x.
b) As a check the concentrations c as found from (*a*) are plotted logarithmically against the logarithm of I, when a straight line should be obtained with $\tan \alpha = n \leqq 1$.

Correction for self-absorption

We shall consider briefly the manner in which self-absorption finds expression in this investigation. As already remarked, one will try to avoid this by measuring lines high up in the energy diagram.

If there is no self-absorption there must be proportionality between the intensity measured and the amount of the element contained in the sample. When the intensity is plotted on a logarithmic scale against the logarithm of the concentration one finds a linear relation with a slope 1:1 (45°). Where there is self-absorption the lines as a rule still show an approximately linear relation between the logarithm of the intensity and the logarithm of the concentration, but with a slope <1, which shows that in that case the simple formula $I = kc$ does not hold and that instead we have the empirical equation

$$I = k'c^p, \text{ where } p < 1.$$

Once the value of p has been determined for a certain line, from tests with known concentrations of a certain element, this can be used for analysing an unknown sample containing that element.

Examples

To give some idea of the degree of accuracy reached, some examples are given of analyses that have actually been carried out.

The first example concerns a chromium-nickel steel. Three entirely independent analyses were carried out, with the following results:

	Fe	Cr	Ni	total
	69.5	19.4	9.7	98.6
	71.3	17.8	9.5	98.6
	71.6	18.1	9.5	98.6
Average	70.6	18.4	9.6	98.6
SD_{rel}	0.8	2.5	1.0	

The figures are percentages by weight; at the bottom of each column the average of the figures in that column is given and underneath that the relative standard deviation from the average as a percentage of the average, thus:

$$SD_{rel} = \frac{100}{y_m}\sqrt{\frac{\sum(y-y_m)^2}{N(N-1)}},$$

where y is the measured value and y_m the average of a column ($N = 3$).

The total time taken for the analysis (carried out by one person) was 2 hours.

The next example relates to a magnet steel. For each element two values are given, each being the average obtained from four spectra recorded on one plate, whilst also the average of these values is given. For comparison the result of a chemical analysis is likewise shown in a separate column and the last column gives the relative percentage difference (Δ) between the chemical and the spectral analyses.

		spectr.	chem.	Δ
Fe	48.7 - 49.0	48.8	48.9	0.2
Co	24.1 - 23.5	23.8	24.4	2.5
Ni	15.2 - 15.4	15.3	15.0	2
Al	8.1 - 8.1	8.1	8.0	1
Cu	2.9 - 2.9	2.9	2.9	0
Ti	0.82 - 0.82	0.82	0.82	0
Si	0.17 - 0.17	0.17	0.16	
Mn	0.115 - 0.11	0.11	?	
	total	100.0	100.2	avge 1

The relative standard deviation from the average for one determination for the various elements is 0.5 to 3% (average 2%). In order to reach a relative standard deviation of 1% four determinations are necessary. Four analyses, if carried out by one person alone, take 3 hours and 20 minutes.

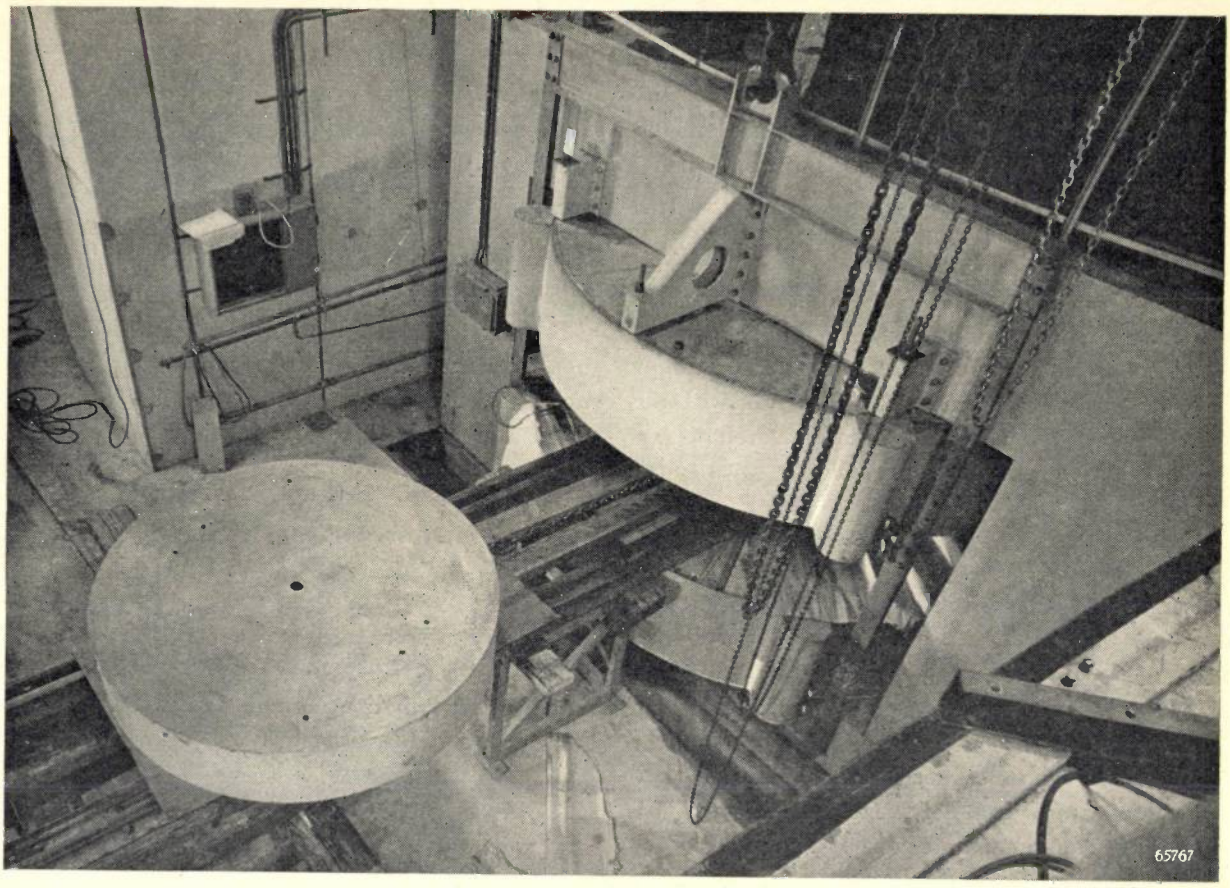
As a third example may be mentioned the case of carbonates of earth alkalies of the probable composition 40% $BaCO_0$, 35% $SrCO_0$ and 25% $CaCO_0$. The relative standard deviation for these three components amounted respectively to 2, 1 and 4%.

Since no great accuracy was required these analyses were carried out in duplicate and took in all 90 minutes.

Finally the spectrochemical analysis of a glass is given, and for comparison the result of a chemical analysis of the same sample.

	Spectrochemical	Chemical
SiO_2	70.8	70.6
Na_2O	0.2 ± 15% rel.	0.16
K_2O	1.1	0.97
Li_2O	0.9	0.87
B_2O_3	26.3	25.8
MgO	0.008	not determined
Fe_2O_3	0.04	0.05
TiO_2	0.02	not determined
As	0.01	not determined
F	not determined	0.52

Summary. As an introduction to a description of the method of spectrochemical analysis applied in the Philips Laboratory at Eindhoven, the ways in which a spectrum can be formed are considered. It is discussed in turn how atoms are excited by electrons of constant energy (experiments by Franck and Hertz), in a cavity with a constant temperature (King's furnace), in a flame, in electric discharges at low pressure (Geissler tube) and in electric discharges at high pressure (spark and arc). It is explained on what grounds preference is given to the D.C. carbon arc. The specimen to be analyzed is completely vaporized in this arc and a constant part of the integrated radiation is recorded in the form of a spectrum on a photographic plate. The qualitative test (identification of the lines) and the quantitative test (measurement of the density of the lines) are performed simultaneously with the aid of an instrument projecting the spectrogram onto a screen. The density to be measured is related to a scale of density marks on white paper (the standard paper density or s.p.d. scale). One of the advantages of this method is that a correction for the background intensity of the spectrum is made automatically. The s.p.d. scale is calibrated once for all with the aid of densities on a photographic plate caused by known intensities. In many cases the intensity proves to be proportional to the amount of the element investigated. In the case of lines showing self-absorption the relation can be represented by a simple empirical equation. The absolute relation between intensity and concentration is established experimentally once for all. Finally some examples of analyses are given relating both to metals and to non-metals (e.g. oxides, glasses, salts), with figures for the accuracy and the duration of the analysis.



THE SYNCHROCYCLOTRON AT AMSTERDAM

III. THE ELECTROMAGNET

by F. A. HEYN*).

621.384.61 : 621.318.3

The huge electromagnet needed for a cyclotron is undoubtedly what in this apparatus appeals most to the imagination of the layman. A favourite demonstration for visitors is to switch on the magnetic field and to let them feel for themselves the great magnetic forces in the cyclotron by giving them an iron rod or a copper ring to hold in their hands.

There is not much to be found in technical literature about the construction of electromagnets, so that it is not merely on account of its imposing dimensions that a more detailed treatment of the construction of a cyclotron magnet seems to be worth while.

Introduction

The most important data of the Philips synchrocyclotron that has been installed in the Institute for Nuclear-physical Research at Amsterdam have been given in the first article of this series [1]). It may be recalled that this apparatus accelerates deuterons until they attain an energy of 28 million electronvolts, and that this is reached by applying an alternating voltage of about 14,000 volts (peak value) between the two dees in the accelerating chamber.

The second article [1]) gave a description of the oscillator, supplying this voltage with a frequency of $\omega/2\pi = 10.7$ Mc/s, and of the modulator, which brings about a periodical reduction of this frequency

*) Professor at the "Technische Hogeschool" (Engineering University) at Delft (Netherlands), formerly of Philips Laboratories at Eindhoven.

[1]) F. A. Heyn, The synchrocyclotron at Amsterdam, I. General review of the installation, II. The oscillator and the modulator, Philips Techn. Rev. **12**, 241-247 and 247-256, 1950/51 (No. 9), further referred to in this article as I and II respectively.

by about 4%, so that the particles spiralizing outward may be kept in resonance with the accelerating voltage in spite of the inevitable decrease of the magnetic flux density towards the edge of the magnetic field and the relativistic increase of mass.

We shall now describe the magnet supplying the powerful field in which the dees are situated and which constrains the accelerated particles to follow circular paths. An attempt will be made to explain on what grounds the magnet has been given its particular shape and dimensions and how these have been determined. After dealing with the construction of the magnetic circuit we shall discuss the construction of the exciting coils and the cooling system.

Fundamental considerations for the design of the magnet

The basic point was the energy to be given to the particles, which had been fixed at 25 million electron volts. For deuterons with this energy the classical formula for kinetic energy $T = \frac{1}{2} mv^2$ (v = velocity, m = mass of the deuterons = $3.35 \times \times 10^{-27}$ kg) can be applied with sufficient accuracy for our purpose, and with the aid of the resonance condition ($\omega = v/\varrho = eB/m$, see I) it follows directly that:

$$T_{\mathrm{MeV}} = \frac{10^{-6} e}{2m} (B\varrho)^2, \quad \ldots . \quad (1)$$

where e is the elementary charge = 1.6×10^{-19} coulomb, ϱ is the radius of the orbit in metres, B the flux density in weber/m^2 at the orbit.

When particles are required with a certain energy, according to the formula the product BR is fixed, R being the radius of the largest orbit described by the particles between the pole faces. From $T = 25$ MeV it follows that $BR = 1.015$ Wb/m. R is always smaller than the radius r of the poles of the magnet, because, as will be explained later, the edge of the magnetic field cannot be used on account of the greatly reduced value of B obtained there. Mostly R is 80 to 90% of r, as was already known from other cyclotron magnets. Taking $R = 0.85\ r$, Br must be 1.2 Wb/m. Actually B is here the flux density on the orbit R, but as a first approximation it may be assumed that the flux density is of the same value in the whole of the space between the magnet poles.

Another important factor for the design of the magnet is the distance d between the poles. This is first of all determined by the height to be given to the dees in order to get the greatest possible beam current (see I). Further, the distance between the dee to which the high voltage is applied (the other dee is earthed) and the covers of the accelerating chamber must not be too small, so as to be sure that there can be no risk of flashover and that also the capacitance will not be too high between this dee and earth (see II). Finally, there must be space enough for the auxiliary apparatus, i.a. for the ion source. Therefore we started by considering a value of $d = 36$ cm.

Having fixed Br and d, in order to form some idea of the magnet required we can calculate what magnetic energy is needed in the air gap. This can be determined with the aid of the formula

$$W_m = \tfrac{1}{2} B H \text{ (air gap volume)}, \quad . \quad . \quad (2)$$

where $H = B/\mu_0$ is the magnetising force [2]). Hence:

$$W_m = \frac{1}{2} \frac{B^2}{\mu_0} \cdot \pi r^2 d = \frac{\pi}{2\mu_0} (Br)^2 d,$$

which, after substituting the values $Br = 1.2$ and $d = 0.36$, yields:

$$W_m \approx 650{,}000 \text{ joules}.$$

What this figure means may be imagined when it is said that the magnetic energy in the air gap of a powerful moving coil loudspeaker is about $^1/_4$ joule.

The problem of how to get a magnetic field with the desired values of Br and d is governed in part by technical considerations but more so by economic considerations.

For various technical reasons a small pole diameter $2r$ should be chosen, thus a high flux density B is required. The dee is then not excessively heavy, this being of importance because its weight has to be carried by a transmission line which can be supported by ceramic insulators only at a considerable distance from the dee (see II); the capacitance of the dee is limited; the dimensions of the (evacuated) accelerating chamber are also kept small, so that it can withstand better the force of the atmospheric pressure, there is less risk of vacuum leaks and less trouble from occluded gases. Against these advantages a high flux density B is uneconomical, as we shall presently see. Only in the case of very large synchrocyclotrons will the compromise between B and r be determined more by the technical considerations, since with a very large value of r the technical difficulties become almost insurmountable.

As regards the economic considerations we have

[2]) The reader is reminded that in the system of rationalized Georgi units used here the permeability of air (or vacuum) $\mu_0 = 4/\pi 10^7$ henry/metre.

to distinguish between the capital outlay and the running expenses. Under the running expenses, apart from the salaries of the operators, it is mainly the power consumption and cooling water that are of importance; in fact, the whole of the energy converted into heat by the dissipative resistance of the exciting coil has to be carried off by a water-cooling installation. Coming under the capital outlay is, among other items, the cost of this cooling installation and that of the converter supplying the direct current required. The smaller the necessary electric energy is kept the lower are these costs. But by far the biggest item in the capital outlay is the cost of the amount of steel [3]) required for the magnetic circuit and the copper for the coil. If the electric power is to be kept small then the copper of the magnetising coil has to be of a large cross-sectional area, thus involving a very large amount of copper, and since the closed magnetic circuit must leave space enough for the coil to be built into it, the more we try to limit the dissipated power the more steel is needed.

Thus it is seen that the power to be chosen must be a compromise between running expenses and capital outlay. In this connection it is very instructive to consider for a moment the possibility, in principle, of reducing the cost of electricity and water entirely to zero, namely by employing a permanent magnet! A simple calculation will show that this "compromise" is certainly not the optimum. Under idealized conditions the following formula holds for a permanent magnet [4]):

$$\tfrac{1}{2}B_{\text{air}} \times H_{\text{air}} \times \text{volume of air gap} = \\ = -\tfrac{1}{2}B_{\text{steel}} \times H_{\text{steel}} \times \text{volume of magnet steel.} \quad (3)$$

On the left we have again the magnetic energy in the air gap, W_m, and on the right that in the magnet steel. The product $B_{\text{steel}} \times H_{\text{steel}}$ depends very much upon the kind of steel and the working point on the B-H curve, which is determined by the form of the circuit and its antecedents. With a modern magnet steel, viz. "Ticonal" 2A (there are still better steels but these are much more expensive), the product can be at most about $-15{,}000$ joules/m^3 (i.e. -1.8×10^6 gauss oersted); taking into account, i.a., the leakage, in practice one may perhaps reckon on a value of $-10{,}000$ joules/m^3. Taking the value of $W_m = 650{,}000$ joules we thus find that the permanent magnet would have to contain roughly 130 m^3, or 1000 tons, of magnet steel! Owing to the expensive alloy components and the complicated manufacture the cost of magnet steel is many times that of "ordinary" steel suitable for an electromagnet, and for the electromagnet of our cyclotron it appeared that we could manage with "only" 200 tons of this steel. Obviously, therefore, even if it had been possible to overcome the inherent technical difficulties [5]), the use of a permanent magnet could not be considered.

Now let us turn back to the question of the power to be chosen for the electromagnet. The necessary compromise between capital outlay and running expenses will, of course, be greatly affected by the local rates charged for electricity and water and by the prevailing prices of the (electrolytic) copper and steel needed. As to the latter costs it may be noted that for a given power it is possible to design constructions using a large amount of steel but requiring comparatively little copper, and vice versa. There is an optimum quantitative ratio of steel to copper, which in turn is affected by the relation between the quality of the steel and its price. Quality is a function of the B-H curve of the steel. The better the quality, the less the quantities of steel and of copper required for a given dissipative power.

Thus it is evident that it is a rather complicated matter to find the construction most favourable from the point of view of economy. In our case the problem was simplified by the fact that a converter was already available with a power of more than 80 kW, which is of the same order as was used for other cyclotrons for about the same particle energy. Further, there was no alternative as regards the quality of the steel, since only one type of steel was available [6]). Thus our problem was reduced to the question of the optimum quantitative ratio of steel to copper for the given power and the given quality

[3]) According to the new nomenclature, all grades of iron now used as constructional material, also the soft "iron" of an electromagnet, have to be classified as steel (the criterion is the occurrence of an α-γ transition, and not, as before, a minimum carbon content).

[4]) See A. Th. van Urk, The use of modern steels for permanent magnets, Philips Techn. Rev. 5, 29-35, 1940. — The minus sign in the equation is due to the fact that in the case of a permanent magnet H_{steel} is always opposite to B_{steel} ("demagnetising force").

[5]) The permanent magnet would have to be magnetised after it has been installed, and this would entail the building of an enormous auxiliary coil and installation for activation which would only be used once. Further, the fact that a permanent magnet cannot be temporarily "switched off" for carrying out work on the accelerating apparatus, etc. would have led to very great complications.

[6]) The properties and the manufacture of this kind of steel were described some time ago in this journal: J. J. Went, Philips Techn. Rev. 10, 246, 1948. It is to be added that further insight subsequently obtained into the specific requirements that have to be answered by the steel for our cyclotron has shown that the steel in question could have been given still more favourable properties (a higher magnetic saturation) without involving any extra cost; the castings had, however, already been made.

of steel, and how the magnetic circuit and the magnetising coil had to be dimensioned so as to get that optimum ratio.

With the converter available an upper limit was set to the current I and the voltage V of the magnetising coil, namely about 175 A and about 500 V. It is important to note that this restriction in the choice of I, for a given power, is no impediment in reaching the optimum ratio mentioned above, since the magnetic field in a given magnetic circuit is determined only by the number of ampere turns, nI, and with a given product nI the volume of copper required for the coil is not primarily influenced by the further choice of n and of I. If a low value is chosen for I (thus n large) then for the given power $W = r_k I^2$ the resistance r_k of the coil can be high, whilst if a high value is chosen for I (n small) r_k must be small, and it appears that for this reason the cross-sectional area of the copper for the large number of turns in the first case may be chosen so much smaller than that required for the few turns in the second case that the total volume of copper (provided the average length per turn is the same) works out equal for both cases.

Even if the said limits in the choice of I and V had not already been fixed by the converter available, from the point of view of dependability we should have been obliged not to choose too high a voltage across the coil, since it is imperative to exclude the risk of breakdown in the coil, having regard to the amount of work that would be involved if this part of the installation should need repairing. It is only with a comparatively low voltage that this dependability can be guaranteed without excessive insulating material being required in the coil, that is without causing the "copper factor" of the coil (percentage of copper in the total volume) to drop too far below 100%.

On the other hand, a much lower voltage and higher current than the given values of about 500 V and 175 A could not be chosen either, because higher currents would lead to rather exceptional constructions of the direct-current generator and, moreover, the correspondingly very small resistance r_k of the coil (even in our case this had to be limited to about 3 ohms) is apt to entail all sorts of constructional difficulties (influence of contact resistances, irregularities in the generation of heat, etc.).

Procedure in designing

General layout of the magnet

The general layout of the electromagnet to be designed follows from some obvious considerations. The magnetic field in the air gap must above all be uniform and symmetrical with respect to the centre plane, so that it is necessary to have two coils, one on either side of the air gap. The two cylindrical poles around which the coils are placed exercise a powerful force upon each other (in our case, as may be calculated, about 2 million newton, corresponding to a weight of about 200 tons); the mechanical distortion must nevertheless be small and in any case symmetrical so as not to affect the field adversely. In particular, the pole faces must be kept exactly parallel. Finally the space between the poles must be easily accessible for mounting the accelerating apparatus proper. The shape of the magnet as drawn in *fig. 1* satisfies all these requirements. In this sketch the magnet has been drawn to scale in the shape that was ultimately chosen and the actual dimensions are given. It was of course the determination of these dimensions that constituted the main task in the designing of the magnet, bearing in mind the aforementioned object of obtaining the desired field with the most favourable compromise as regards the volume of copper and the volume of steel needed.

The calculations for the magnet are based upon the equation:

$$nI = \oint H \, dl, \qquad \ldots\ldots \quad (4)$$

according to which the magnetomotive force, i.e. the line integral of the magnetising force along a line of magnetic force (or along an arbitrary closed line enclosed by all the turns) is equal to the number of ampere-turns required to produce the magnetising force. The integral can be split up into one part belonging to the path that has to be traversed in the course of the integration in air — in our case the pole distance d — and two other parts belonging to the path of integration in the two poles and to that in the yoke closing the circuit, thus:

$$nI = (\int H \, dl)_{\text{air}} + (\int H \, dl)_{\text{poles}} + (\int H \, dl)_{\text{yoke}}. \quad (5)$$

The idealized case

The magnetising force in the air gap being $H = B/\mu_0$, for the air term the number of ampereturns required is:

$$(nI)_{\text{air}} = B \, d/\mu_0 \qquad \ldots \quad (6)$$

Since the air term — and here we are anticipating what follows — will form 70 to 80% of the total (this percentage is called the "efficiency" of the magnet), for a first approximation we may ignore entirely the steel terms (taking the reluctance of the steel circuit to be zero) and ask what value

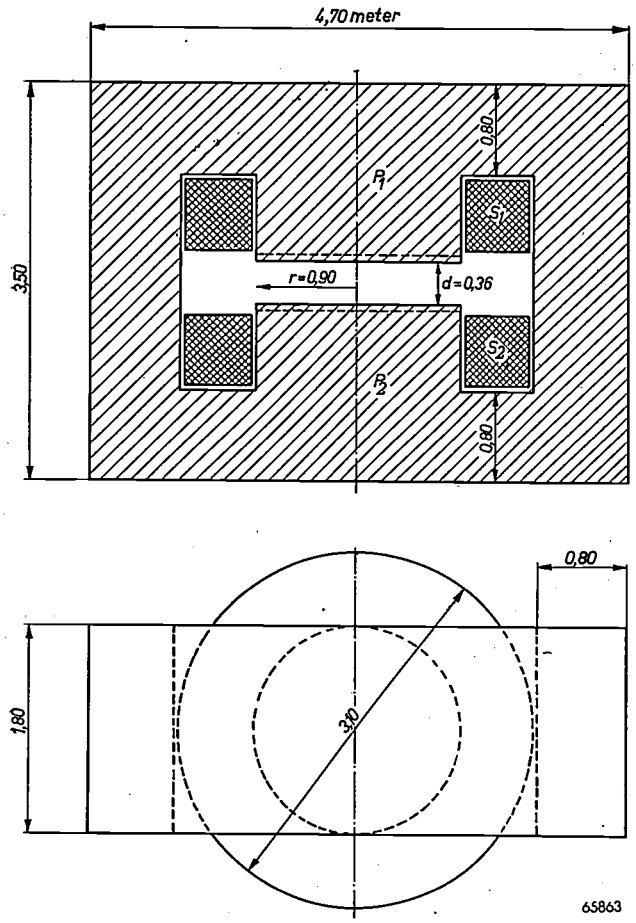


Fig. 1. Form of the electromagnet of a synchrocyclotron, in cross section and top view. Around each of the poles P_1, P_2 is an annular coil (S_1, S_2).

The space between the poles is actually subdivided into three parts. The accelerating chamber is placed between the poles with an air gap of 2 cm on top and underneath; the thick steel covers at top and bottom of this chamber have to be reckoned as belonging to the poles. For the sake of simplicity in the drawing the three air gaps have been combined into one; where mention is made of the "pole distance" in the text this is to be taken as meaning the length of this resulting air gap.

of B, or, since the product $Br = 1.2$ Wb/m is fixed right from the outset, what pole radius r will answer the most favourable compromise to which we have repeatedly referred.

It can now be seen why a compromise is necessary, from what follows. The larger the pole radius r, the larger is the average length of copper per turn, but owing to the reduction of B the necessary $(nI)_{\text{air}}$ is smaller (eq.(6)). As a simple calculation will show, the amount of copper initially required diminishes with increasing r and above a certain value of r — in our case already above $r \approx 1.2$ m — becomes independent of r (and also practically independent of the, still variable, pole length, which is equal to the height of the coil; see fig. 1). At the same time however — assuming, in the first instance, that the cross section of the circuit is of the same size everywhere and that the circuit lies as close as possible round the coil — the quantity of steel required rapidly increases with the value of r. When r is very large a constant amount of copper but more and more steel is therefore needed, whilst with a very small r less and less steel suffices (in this approximation) but the quantity of copper required rapidly increases owing to the large value of B. Taking into account the prices of copper and steel we found for the total costs of materials a flat minimum between $r \approx 0.6$ and 1.0 metre, that is to say with B between 2.0 and 1.2 Wb/m². The number of ampere-turns then needed for the air term is between 600,000 and 350,000, and thus the optimum in question lies within this range — still ignoring the reluctance of the steel.

In view of the technical considerations mentioned in the beginning of this article we should like to choose the smallest r within the optimum range ($r = 0.6$ m), but when we come to take into account the reluctance it is just on this low side that the range narrows down considerably. The fact is that this reluctance makes it necessary to have more ampere-turns, thus additional copper, and the more so the greater the flux density in the steel, i.e. the smaller the pole radius chosen; consequently with smaller r the amount of copper rises more steeply than in the ideal case. The question is to what value of r does the optimum range now still extend?

It is important to realize that this question cannot be answered exactly, for the following reasons.

Difficulties arising when taking into account the reluctance

The answer to this question amounts to a calculation of the two steel terms in eq. (5). For this we have to use first the condition that the magnetic flux — disregarding for a moment the leakage — must be of the same value in any cross section (surface A) of the circuit, thus

$$B_{\text{steel}} \cdot A_{\text{steel}} = \text{constant} = B_{\text{air}} \cdot \pi r^2 \quad ; \quad . \quad (7)$$

and, secondly, the relation between the flux density B and the magnetising force H in the steel, which is given by:

$$B = \mu_0 H + J(H) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where J is the intensity of magnetisation, which depends upon H according to a function empirically determined for each type of steel.

From these equations it is possible, in principle, to determine the value of H in any point of a chosen circuit, and thus the line integral (eq.(4)) for differ-

ent values of the pole radius can be calculated. Actually, however, there is a relatively large uncertainty about this calculation. In the first place equation (7) holds only to a very rough approximation: near the free ends of the poles a considerable part of the magnetic flux leaves the steel laterally and the lines of force join outside the circuit. This leakage, which occurs at any flux density and depends upon the dimensioning, can hardly be calculated at all and has to be estimated or found empirically (see, e.g., the article quoted in footnote [4]). In the second place the intensity of magnetisation J as a function of H shows the well-known phenomenon of saturation. With high values of H, where the maximum possible magnetization J_s is approached and thus the J-H curve becomes flat, small variations in B, caused for instance by a small variation of the respective cross section A in the experimental design, involve great changes in the magnetising force H required, and of course this renders the calculation hazardous. Furthermore, near the corners of the steel circuit (fig. 1) and particularly near the transition from the yoke to the round pole, the flux density B is by no means constant over the whole cross section of steel, and we do not know exactly what "effective" value of B one has to use here. Moreover the extent to which the point of saturation of the steel is approached also affects the stray field — a larger percentage of the lines of force leave the steel the more this becomes saturated — and this influence can be still less exactly calculated. Finally, the trend of the J-H curve near the point of saturation is highly dependent on small variations in the manufacture of the steel; differences in this respect were in fact observed in some of the blocks of steel from which our magnet was built up.

Closer determination of the pole radius to be chosen

Owing to the uncertainties mentioned it is not possible to predict exactly how the total cost of materials will work out as a function of the pole radius r, but nevertheless we can arrive at a closer determination of the minimum radius required by the following simple reasoning.

Once the intensity of magnetisation of the steel has reached its maximum value J_s the flux density B can be increased further only by raising the term $\mu_0 H$ in eq. (8), just as is the case in the air gap. Then the steel may be said "to behave as air". The corresponding part of the line integral, $\int H_{\text{steel}}\, dl$, which is always very small for small flux densities, then rapidly increases and consequently a large number of ampere-turns are needed for this part. To avoid this the construction has to be such that the magnetic saturation J_s is nowhere too closely approached in the steel.

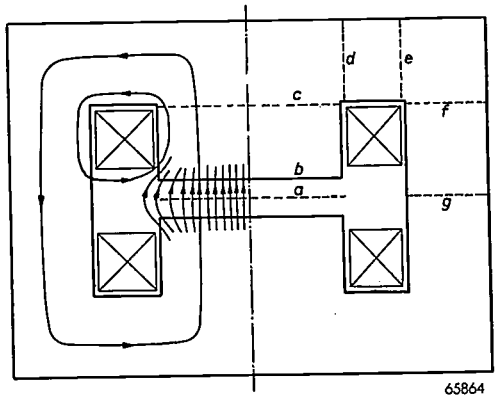


Fig. 2. The flux density considered more closely in various cross sections (a-g) of the steel circuit. The lines of force in the left half of the diagram are indicated very schematically.

In this respect the cross section c in *fig. 2* was to be considered as the most dangerous part. It is easy to predict that the most dangerous place will anyhow lie in the poles: any increase of the cross section — the means of reducing the flux density and thus J — works out more expensive in the poles than elsewhere in the circuit; of course thicker poles require more steel, but they also cost more copper because the average diameter of the turns is larger, and so still more steel is needed because the increased outer diameter of the coil requires a greater length in the horizontal parts of the yoke. In all the existing cyclotron magnets the flux density is indeed greatest in the poles, mostly at or near c. So we could take as a starting point the maximum permissible value of J for that place, which for the steel that we had available, was 2.0 Wb/m^2. According to the J-H curve this value is reached when $\mu_0 H = 0.1$ Wb/m^2, thus the flux density in the cross section c becomes $B_c \approx 2.1$ Wb/m^2.

In the pole faces, cross section b, and in the air gap, cross section a in fig. 2, the flux is much less than in c, owing to the leakage. From trials with models and from the data of existing cyclotron magnets it was known that the ratio of the flux values in c and a — the leakage coefficient — could be taken at about 1.5. Thus, since the cross sections a and c in fig. 2 are of equal size, it follows that with the steel used the flux density in the gap could not be chosen greater than $B_a = 1.4$ Wb/m^2. So as to allow for some reserve in case in some blocks of steel J_s should turn out to be lower, we chose $B_a = 1.35$ Wb/m^2 (giving for $B_c = 2.06$ Wb/m^2 and for $\mu_0 H_c = 0.07$ Wb/m^2). From the prescribed

value of the product $Br = 1.2$ Wb/m in the air gap it follows that the radius of the poles has to be $r = 0.9$ m (rounded off).

If a larger value were chosen for r the steell terms in eq. (5) would, it is true, be reduced somewhat, but then, as we have seen, the cost of material for the so much larger air term would remain practically the same up to $r = 1.0$ m and then even increase.

Other dimensions of the circuit

The experiments just referred to were carried out with a simple model built to a scale of 1 : 20. This model and a second one made later to the same scale but in a modified form provided us with the necessary data regarding the leakage, etc. which enabled us to make a choice not only of the pole radius but also of other details in the shape of the circuit.

In the first model various cross sections of the yoke, viz. at *d*, *e* and *f* (in every case it is the total cross section of the two "parallel connected" branches that is referred to here), were made of the same size as the cross section *c*. Measurements showed that the flux in these cross sections was practically equal to that in *c*. In the second model (and in the actual magnet) the yoke cross sections mentioned were made larger than the pole cross section *c*, namely 13% larger, as determined from some experimental calculations. It is true that this costs more steel, but then this is amply compensated by the saving in ampere-turns and thus in copper; a reduction of the flux density by 13%, from 2.06 to 1.825 Wb/m², is accompanied by a much greater reduction of $\mu_0 H$, namely from 0.07 to 0.025 Wb/m², and this reduction is all to the good considering the great length of the path of integration in the yoke!

As regards the shape of the coil cross section, that is to say the ratio of the height *h* and the thickness *D*, this could be decided before the pole radius had been definitely fixed, by virtue of the following considerations. The greater the height *h* chosen the longer the poles become and thus the more steel and — what is worse — the more ampere-turns are needed, this being partly due to the longer path of integration with high magnetising force and partly on account of the greater leakage, which reduces the "efficiency" of the circuit. Moreover, more steel is then also needed for the vertical parts of the yoke, which, together with the poles, have to be longer; this steel cannot be recovered by placing the vertical parts closer together (as would seem to be possible since the coil thickness *D* becomes smaller with the greater height of the coil), for the efficiency would then be still further reduced owing to a larger percentage of the lines of force passing direct from the poles to these parts of the yoke. On the other hand, the smaller *h* is chosen the larger becomes the average radius of the turns placed one over the other (for a given number of ampere-turns). This adverse effect very soon exceeds the saving in ampere-turns obtained with shorter poles, and also the resultant saving in steel is balanced by the extra steel required for the longer horizontal parts of the yoke. Therefore neither very large nor very small coil heights are desirable, and in fact a roughly square coil cross section is the best. Small deviations from this form make little difference in the total cost of material, as proved by experimental calculations. Incidentially it should be added that in these calculations it is necessary to take into account the copper factor of the coil, which differs somewhat for different shapes.

Balance of the ampere-turns

Thus we have sketched in broad outline the considerations underlying the determination of the dimensions of the magnetic circuit. It was on this basis that the second model was built and it was then possible to calculate roughly the number of ampere-turns that each part of the model would require. *Table I* gives the result of the calculation, all figures being converted to the dimensions of the actual magnet, for the desired flux density $B_a = 1.35$ Wb/m² in the air gap and also for the somewhat larger value $B_a = 1.40$ Wb/m². By way of explanation it has to be added that in the calculation the high flux density $B = 2.06$ Wb/m² of cross section *c* was taken for only one-third of the pole length. At the pole face, cross section *b*, according to the trial model the flux density is already 20% less than that at *c* owing to the leakage; here, therefore, the magnetising force in the steel is practically negligible. From the measurements of the flux density taken at various places between *c* and *b* in the first model the magnetising force in the pole was assumed to be constant ($B = 55{,}800$ A/m) over one-third of the length, then dropping linearly to practically 0 at the end, so that for two-thirds of the length the average value $H = 27{,}900$ A/m could be taken. The inaccuracy involved in this assumption is quite negligible compared with the sources of uncertainty contained in the calculation itself, to which ample attention has been devoted in the foregoing pages.

The efficiency that could be expected from this

Table I. Balance of ampere-turns required for the second model of the magnet (converted to the actual dimensions) according to data in respect to leakage obtained, i.a., from the first model.

Part of circuit	Length in m	B_a = 1.35 Wb/m²: B Wb/m²	$\mu_0 H$ Wb/m²	H A/m	nI A	%	B_a = 1.40 Wb/m²: B Wb/m²	$\mu_0 H$ Wb/m²	H A/m	nI A	%
air gap	0.36	1.35	1.35	1 075,000	387,000	76	1.40	1.40	1 115,000	400,000	70
$^2/_3$ pole	1.03	—	0.035	27,900	57,000	11	—	0.05	40,000	82,000	14.5
$^1/_3$ pole	0.513	2.06	0.07	55,800			2.10	0.1	80,000		
yoke	5.5	1.825	0.015	12,000	66,000	13	1.86	0.02	16,000	88,000	15.5
Total					510,000	100				570,000	100

balance of the ampere-turns, viz. 76%, may be said to be quite satisfactory.

Measurements taken with the second model showed that the assumptions previously made in regard to the leakage agreed well with the actual values. Thus it was not surprising that the calculated number of ampere-turns was found to agree fairly well with the number required in the actual model in order to get the desired flux density in the air gap. The figures are given in *table II*. The model was found to require slightly fewer ampere-turns than was expected.

Table II. The total number of ampere-turns required to obtain the flux density B_a in the air gap. The figures in parentheses denote the efficiency.

B_a Wb/m²	nI calculated	nI measured on second model (converted to actual dimensions)	nI measured on the magnet as constructed
1.35	510,000 (76%)	470,000 (80%)	497,000 (78%)
1.40	570,000 (70%)	527,000 (74%)	548,000 (73%)

From the measurements with the second model it further appeared that 6% of the flux encompasses the individual coils (see the lines of force sketched in fig. 2; the flux through g is 6% less than that through f). This is a considerable amount but not enough to warrant any appreciable increase of the distance between the vertical parts of the yoke.

Now we can pass on from the design to the construction of the magnet and the magnetising force obtained.

A few points in the foregoing rough outline of the design procedure call for further explanation. In particular there is the question of the definite choice of r when taking into account the reluctance of the steel. The pole radius was chosen large enough for the intensity of magnetisation in the dangerous cross section c to reach just the maximum permissible value. Then it was assumed, almost as a matter of course, that the pole is just as thick at the pole face (cross section b) as in c, but this is not a necessity, for the poles could be made conical. In b the flux is 20% less than in c owing to the leakage; with a conical shape of the pole it is therefore possible for the intensity of magnetisation, though relatively increased in b, to be still below the maximum through the whole of the pole. The question therefore arises whether a conical shape, given the same flux density B_c in the cross section c (radius r_c), is advantageous. Suppose for a moment that we reduce the cross section b (radius r), which originally had the value of $r = 0.9$ metre, and that there is then no change in the leakage factors 1.25 and 1.5, which indicate how much the flux in cross section c of the steel is greater than that in b and a respectively. In that case $B_c r_c^2 = 1.25$, $B_b r^2 = 1.5\ B_a r^2$. For the desired energy of the particles $B_a r$ is already fixed, thus $B_c r_c^2 = \text{const.} \times r$, and so, with B_c given, the radius r_c (inner radius of the coil!) could be reduced in proportion to $\sqrt{r}$. In this way we could go to about $r = 0.6$ metre (this limit of the optimum range of r is in fact not appreciably affected by a conically shaped pole); r_c may then be reduced to about 0.75 m. This would indeed mean a considerable saving in steel and copper. Unfortunately the assumption made above is not quite correct: with the conical shape the flux in b and a will be relatively smaller because with this shape of pole — and also on account of the greater average flux density in the pole — there is increased leakage. So little remains of the expected advantage that it does not justify the complications of a construction with conical poles.

What may yield a real advantage is to depart from another tacit assumption, viz. that the whole of the magnetic circuit should be made of one and the same kind of steel. For the most critical parts of the poles, at the cross section c, we could use a more expensive kind of steel, alloyed with cobalt, which increases the magnetic saturation of the steel. This would have meant an appreciable saving in ampere-turns, which might amply have compensated the additional cost of the steel, but, as already stated, at that time there was only one kind of steel available to us.

Of course it is still more advantageous to use a better grade of steel for the whole of the magnet, provided it is not much more expensive. Measurements taken with a more recent model practically identical to the second one referred to above but made with a grade of steel which at present could be considered for the purpose, with a slightly higher saturation, showed that for $B = 1.4$ Wb/m² in the air gap no more than 484,000 ampere-turns will be needed (efficiency 84%!).

The magnet as constructed

Apart from small deviations from the assumed leakage and other inaccuracies in the calculation, the fact that for the second model fewer ampere-turns were needed than had been calculated is presumably due also to the chance circumstance that the steel, especially that in the poles, happened to be of a somewhat better quality. Of course the quality of the steel might also have turned out rather less satisfactory than anticipated, particularly since the annealing of large blocks is different from that of small ones, and this possibility had to be considered as well. Further, defects in the finishing may result in the occurrence of clearances between the blocks of which the magnetic circuit is composed, which raise the reluctance of the circuit. In order to meet such contingencies, in the dimensioning of the actual magnet the copper volume was chosen for a 5% larger number of ampere-turns than calculated from the second model. Afterwards this was indeed found to be quite necessary, as may be seen from the last column of table II.

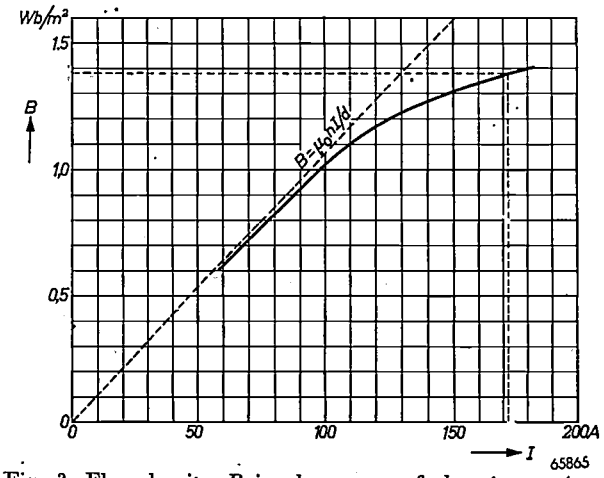


Fig. 3. Flux density B in the centre of the air gap (cross section a in fig. 2) as a function of the excitation current I, as measured on the magnet constructed. The broken straight line, $B = \mu_0 nI/d$ (eq. (6)), is what would be obtained if the reluctance (line integral $\int H\,dl$) of the steel were negligible.

Fig. 3 represents the flux density B obtained in the air gap of the actual magnet (thus really B_a) as a function of the excitation current I. The coils were made with a total number of 3048 turns, with which the flux density $B = 1.35$ Wb/m² is reached at a current $I = 163$ A, so that $nI = 497{,}000$. The measures that have to be taken to get a suitable radial field distribution in the air gap (the "shimming"; see below) result in a reduction of the flux density in the axis. This effect, however, was more than compensated when it was found possible to increase the excitation current without the coils becoming too hot (with the cooling capacity available). Thus we can work quite well with an axial flux density $B = 1.38$ Wb/m². This value was therefore taken as basis for fixing the oscillator frequency, which of course has to be so chosen that the deuterons are brought into resonance with the electric field: $\omega = eB/m$. With the stronger excitation a particle energy of 28 MeV could be reached, instead of the 25 MeV for which the apparatus had been designed (see eq. (1)).

For the resonance to be maintained it is essential that B and thus the excitation current should be kept highly constant. The stabilizing device that was made for this will, it is hoped, be described in another article, as also the safety installation necessary in the event of a sudden failure of the excitation current; unless special precautions are taken, the rapid disappearance of the enormous magnetic flux might give rise to tremendous over-voltages in the coils.

The shimming of the field

For the cyclotron or synchrocyclotron to function properly it is necessary to focus the beam of particles in the centre plane of the air gap. This has already been fully discussed in this Review [7]). For the efficient focusing the magnetic field below and above the centre plane must have a radial component such that the pattern of the lines of force is barrel-shaped. Such a field is obtained automatically as a result of the leakage; the vertical component B of the flux density thereby decreases with increasing radius ϱ — see the dotted curve in *fig. 4a*. Normally, according to this curve, when ϱ is small the relative gradient $(1/B)\mathrm{d}B/\mathrm{d}\varrho$, which may serve as a measure of focusing, is very small, in fact too small for the desired focusing action, whilst with a large radius ϱ it rapidly increases, so much so that an inconveniently large frequency variation of the oscillator would be needed to keep up with the decreasing angular velocity of the particles reaching this zone, whilst we should still have to be satisfied with a relatively small maximum radius R of the orbit (low final particle energy). By placing on the pole faces suitably shaped steel correcting plates and rings, called "shims", the field variation $B(\varrho)$ can be so influenced as to make the gradient answer our purpose better. We aimed at getting an average gradient of 0.02% per cm and keeping this as constant as possible from the axis to the edge of the poles. The shims

[7]) W. de Groot, Cyclotron and synchrocyclotron, Philips Techn. Rev. 12, 65-72, 1950/51 (No. 3).

employed to attain this object, the dimensions of which had to be determined empirically, are represented in fig. 4*c* (not drawn to scale). The result of this shimming is indicated by the fully-drawn lines in figs 4*a* and *b*, from which it is seen that up to a radius of 75 cm the gradient is practically constant, after which it rises very steeply. This means that, when the target is placed in positions corresponding to successively larger values of ϱ, from $\varrho \approx 75$ cm onwards the intensity of the beam will rapidly decrease. In fig. 4*b* the measured beam current i has also been plotted as a function of ϱ, from which it is seen that the maximum useful radius can be taken as $R = 74$ cm. Thus R is 86.5 % of the radius $r = 90$ cm of the poles, and this agrees well with the supposition made in the beginning of this article ($R = 0.85\ r$). With $R = 78$ cm B is about 1.7% smaller than on the axis, viz., 1.355 Wb/m², so that the maximum particle energy $T = 28$ MeV is in fact reached. Of course the limit $R = 78$ cm is not sharply fixed. If we are satisfied with a smaller beam intensity a somewhat greater energy than 28 MeV can be reached.

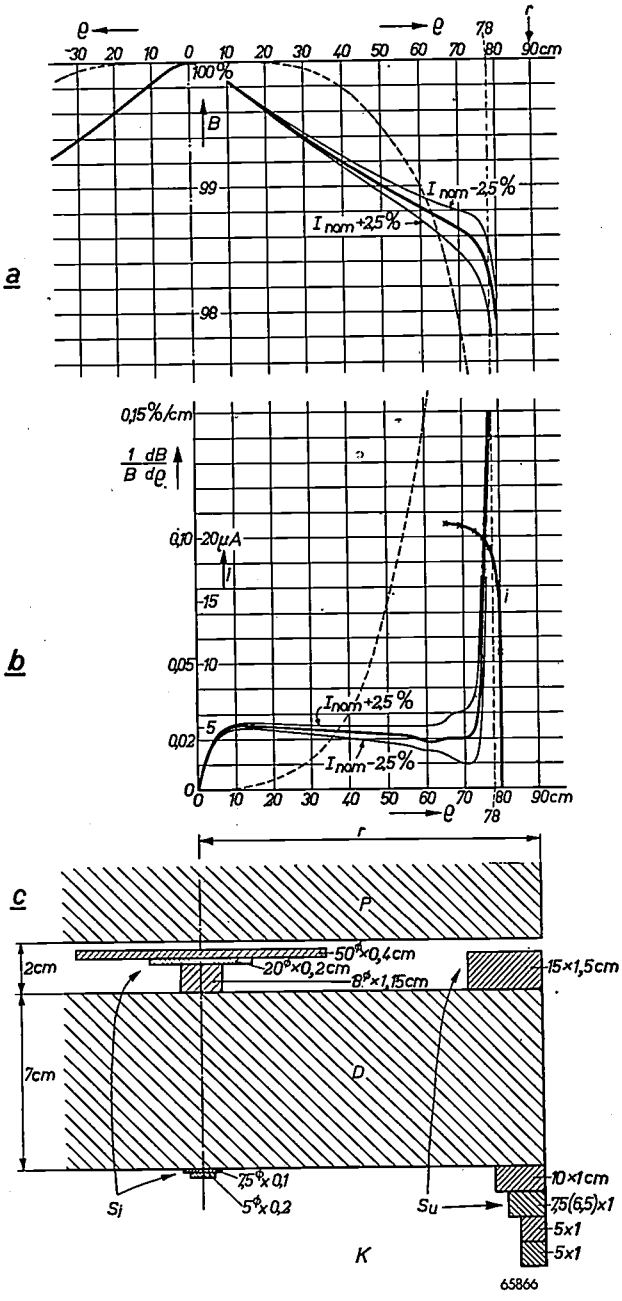


Fig. 4. The shimming of the magnetic field for focusing the particles in the centre plane of the air gap.
a) Variation of the vertical component (as percentage) of the flux density B in the centre plane as a function of the distance ϱ to the axis. Broken line: before shimming. Full line: after shimming. Lightly drawn lines: in the case of the excitation being 2.5% stronger or weaker than normal.
b) The gradient $(1/B)\mathrm{d}B/\mathrm{d}\varrho$ of the curves in (*a*). The measured beam current i (in micro-amperes) at the target is likewise plotted as a function of the distance ϱ from the target to the axis. From $\varrho = 78$ cm onwards the beam intensity drops rapidly. Here the flux density is $B = 1.355$ Wb/m², from which follows a maximum deuteron energy of 28 MeV.
c) Cross section of the pole face with the steel shims used for correction. (The other pole is of the same configuration.) For the sake of clarity the heights are drawn to a 10 times larger scale than the widths. P is the pole proper, D the cover of the accelerating chamber K (see fig. 1) which, as far as the thickness of steel is concerned, has to be added to the pole proper; the air gap of 2 cm — costing roughly 40,000 ampere-turns! — was necessary to allow of the accelerating chamber being put into position.
By the combination of central shims s_i the gradient is made to differ perceptibly from zero at a few cm from the axis, while already at 5 cm it reaches the value of 0.02%/cm, which is kept up to about 75 cm from the axis thanks to the shims s_u.

The useful radius R could be slightly increased by making the shims thicker, so that the gradient is kept small up to a larger radius, but in our case this was not possible owing to the space required for the construction of the accelerating chamber, the dees and the accessories.

Owing to the varying degree of saturation in some of the shims the field variation $B(\varrho)$ is dependent upon the excitation current I, as is to be seen from the curves in fig. 4*a*. If I is too small the gradient first diminishes rather steeply towards the edge (fig. 4*b*), possibly even becoming negative, and this would immediately lead to defocusing, causing the beam to disappear. The excitation current must therefore be carefully adjusted to the right value. There need be no anxiety about possible variations while working, since, as already stated, in order to maintain the resonance the current I must anyhow be kept highly constant (to within 0.06%).

It is interesting to consider the radius where the expression $(\varrho/B)\mathrm{d}B/\mathrm{d}\varrho$ reaches the value 0.2. Here the vertical oscillations of the circling particles fall into resonance with the radial oscillations, as a result of which — if the zone with the critical value is not quickly passed — defocusing occurs (see the article quoted in footnote [7]). In our case, at $\varrho = 87$ cm the gradient $(1/B)\mathrm{d}B/\mathrm{d}\varrho \approx 0.15\%$/cm, so the expression referred to above amounts to about 0.12. Roughly speaking, it may therefore be said that in our case the limit set by this defocusing effect upon the energy attainable coincides practically with the limit set by the frequency variation of the modulator being insufficient to compensate the steep decline of B. This coincidence

of the two limits ensures that full use is made both of the magnetic field and of the modulator.

Unless counter-measures are taken, the presence of the vertical parts of the yoke causes an azimuthal variation of the field in the gap. This asymmetry can be sufficiently corrected by making the shim ring of 7.5 cm nominal width a little narrower in some places, to as much as 6.5 cm.

Mechanical construction of the magnetic circuit

For the magnet of the dimensions predicted, in this way a total weight of 200 tons of steel was needed. The steel could be supplied only in blocks of at most 15 tons. The yoke was therefore built up of 2 × 4 horizontal and 2 × 4 vertical blocks. The four bottommost blocks were laid on steel girders cast in a concrete foundation, then being levelled off precisely horizontal and at equal height with the aid of packing plates. The blocks for the vertical parts of the yoke were set up on these horizontal blocks and the uppermost blocks were then laid across them (see *fig.* 5). No bolts were used anywhere for fixing these blocks together, their weight being sufficient to keep them in place; there is no need to draw the parts together so as to remove any air gaps between them (see below), since the magnetic forces do this better than is possible by bolting the parts together. Each of the poles was divided into two disks, so that the coils could be put into place (see below). The two disks of the bottom pole were secured with three pins to prevent lateral displacement. The disks of the upper pole were bolted onto the top blocks of the yoke. Although the poles are attracted towards each other by a force of about 200 tons, the bolts need only be calculated to carry the weight of the upper pole, since the magnetic flux determining the force bearing upon a cross section of the circuit is much greater in the cross section *c* (fig. 2) than at the pole faces, cross section *b*. After the excitation current has been switched on the upper pole is therefore drawn up against the top beams of the yoke with a force greater even than 200 tons!

Owing to the force of attraction between the poles, which of course is transmitted as a mutual attraction to the bottom and top beams, the beams tend to bend slightly inward and it would be logical to fear that the vertical beams would thereby be forced outward. The flux through the cross section *f* (fig. 2) however is so great that there the beams are kept

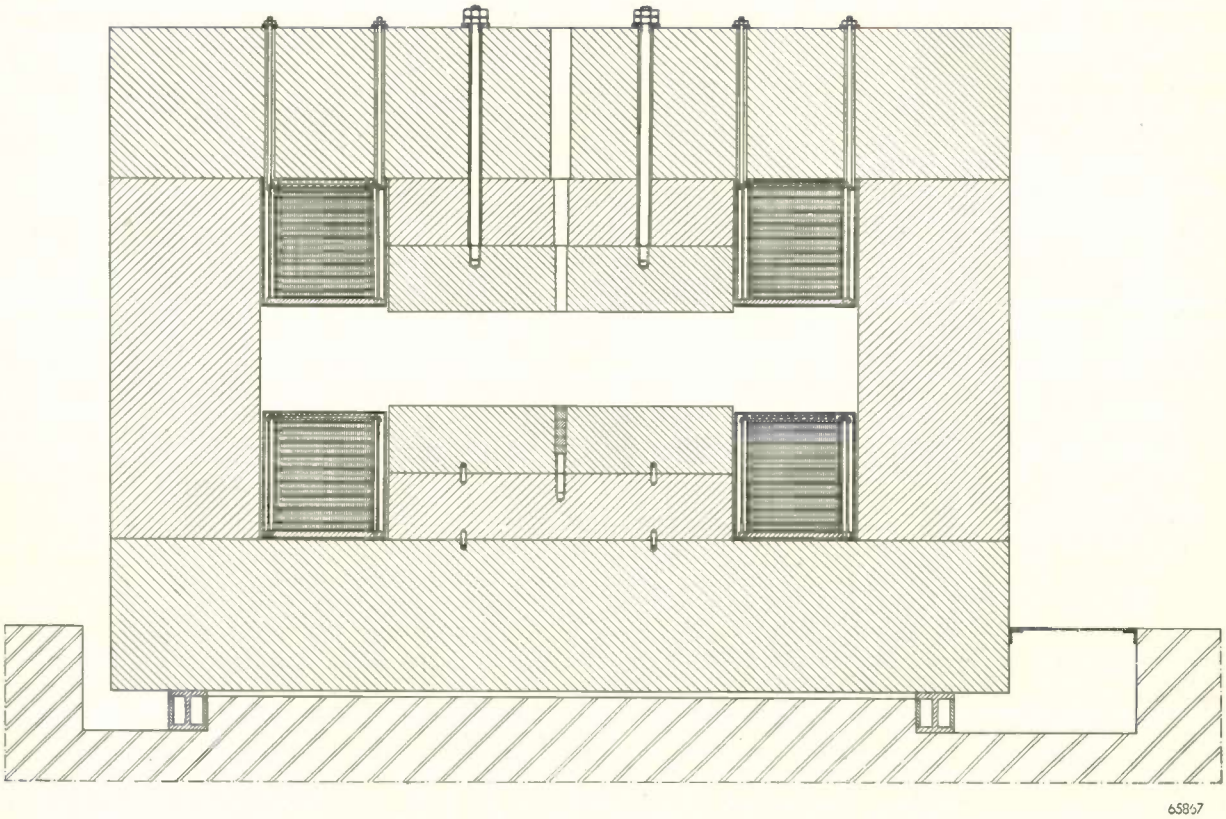


Fig. 5. The yoke drawn here in longitudinal cross section is built up from 2 × 4 horizontal and 2 × 4 vertical steel blocks. Each pole consists of two disks. The aperture in the upper pole, which, like that in the lower pole, was partly filled up before the installation was started up, is necessary for mounting the upper pole and the accelerating chamber.

well together: the deformation of the cross beams is such that their centre line has an inflexion point between the poles and the uprights of the yoke.

Construction of the coils

From the number of ampere-turns $nI = 497{,}000$ and the data of the generator already given in the foregoing it was possible to derive the number of turns required and the cross section needed for the copper. Of course the calculation had to be based on the specific resistance of copper at the working temperature (about 45 °C). Thus we arrived at the number of over 3000 turns already mentioned.

To be able to wind the turns easily and with little force strip copper was used with a section of 0.5 cm × 3 cm. This could be supplied only in rolls of 60 m. For the total length of 24 km (weight about 30 tons) 400 joints had therefore to be brazed. The brazing was done as far as possible mechanically. The joints were made with silver and tested one by one under a tensile force of 500 kg to preclude any risk of breakage after the coil had been made. The strip was bound into what might be called "pancakes" of about 100 turns each on a horizontal revolving table, with two strips of specially strong paper each 250 microns thick interwound as insulation; see *fig. 6*.

Each of the two coils is built up from 16 of these "pancakes" stacked one upon the other. These lie in pairs. The two pancakes of each pair are insulated from each other by two layers of hard paper each 1 mm thick. The pairs are insulated from each other by pieces of resin-bonded fabric 10 mm thick so arranged as to leave a zig-zag path free for the passage of the oil for the cooling (see fig. 8). The pancakes and intermediate plates of each coil were pressed together with the aid of a number of steel plates and tie rods, thus forming a compact ring (*fig. 7*). The bottom coil was placed around the bottom pole, resting on the bottom beams of the yoke and some supporting columns. The top coil was suspended from 16 steel rods screwed into the (locally thickened) clamping plates and passing through notches made in the top beams of the yoke and through six brackets screwed onto the yoke. These rods do not have to carry more than the weight of the coil, since the ponderomotive forces of the magnetic field produced outside the steel by the leakage field force the top coil upwards and the bottom coil downwards.

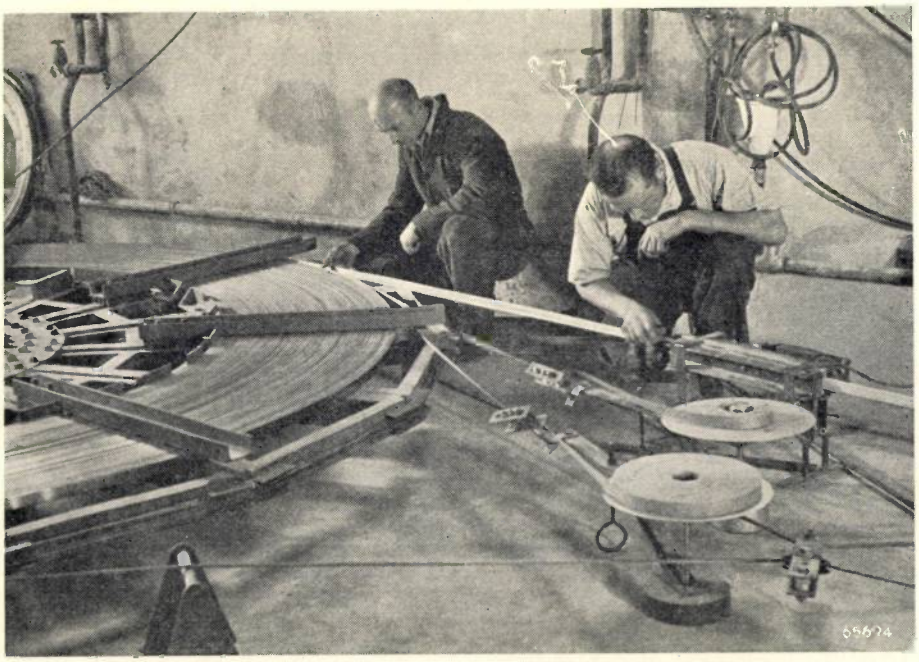


Fig. 6. Winding one of the 2 × 16 "pancakes" from which the excitation coil was built up. The pancake is made by winding a piece of copper strip about 750 m long (cross section 3 cm × 0.5 cm) on a horizontal revolving table, with two layers of paper, fed from two rolls on the right in the foreground, as insulation between the turns of copper. The table was revolved by winding a steel cable round its periphery (seen in the foreground) and then unwinding this cable with an electric winch. The copper strip was kept taut by feeding it through a clamping block with oak facing (500 kg tensile force while winding), seen behind the rolls of paper.

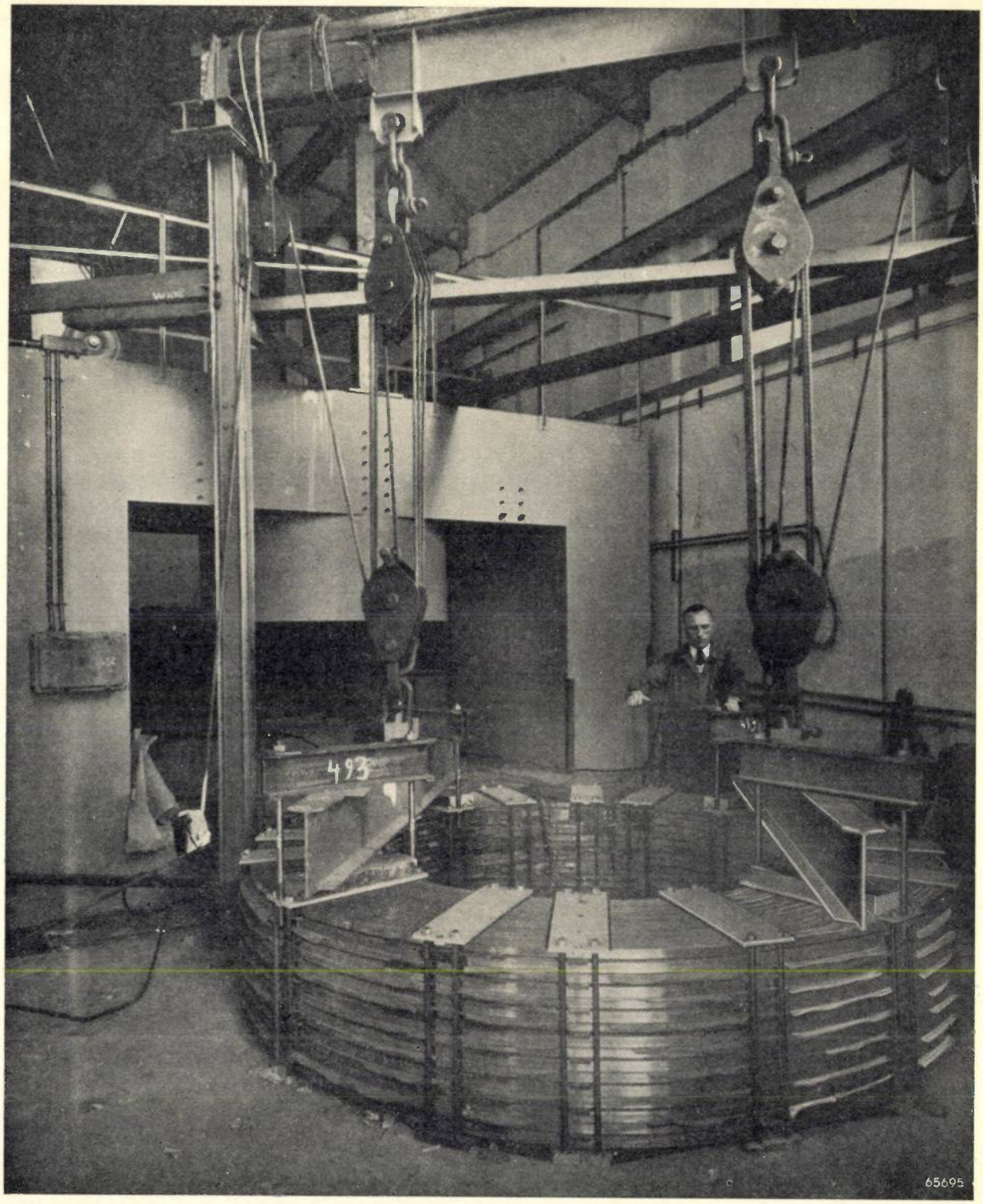


Fig. 7. Completely assembled coil ready to be placed in the tank (fig. 9) and then to be hoisted and put in place round the upper pole of the magnet. See also frontispiece where the two coils have already been placed in position, only the top half of the lower pole remaining to be slid into position and then lowered into place.

Cooling of the coils

The heat (over 80 kW) generated in the copper is carried off by a stream of water, but for the sake of the insulation of the bare windings circulating oil is used as intermediate link. Each coil lies in an annular tank fitting loosely around it, oil being admitted into the tank at two diametrically opposite points. After flowing through the zig-zag channels left between the pancakes the oil runs out through two overflows situated between the inlets

(*fig. 8*). The oil in the tanks is not brought under pressure, so as to preclude as far as possible any chance of leakage; it is kept in motion by a difference in height of only a few centimetres between the levels of the inlets and outlets in the tanks. Packing posts of wood placed at intervals round the periphery of the coil inside the tank serve to prevent any "working" of the windings due to ponderomotive forces and at the same time prevent the oil from flowing in any direction other than through the channels. The tanks, which are placed round the coils before the latter are put into place, are not subject to any mechanical load whatever, so that again any risk of oil leaking out through possible capillary cracks is minimized, while nevertheless the steel wall of the tank can be so thin as not to absorb any magnetic flux worth mentioning (*fig. 9*). The tanks are covered with a sheet of tin-plated iron slightly corrugated to prevent mechanical stresses, the cover being soldered on with tin. The hot oil flowing out of the tanks is led into a collecting tank underneath the floor, from which it is pumped through 10 parallel copper pipes around which cooling water is circulated in steel tubes (*fig. 10*). After being cooled the oil flows back into the coil tanks. The consumption of water is 100 litres/min. The oil flows through the cooler and the coil tanks at a rate of 300 litres/min. Good thermal transmission is ensured by arranging for a turbulent flow of both media in the cooler. The difference in temperature between the oil flowing into the coil tanks and that flowing out of them is only a few degrees, since all the oil channels run parallel and the oil traverses only a quarter of the circumference of the coil. It was necessary to have such a short passage for the oil because otherwise we could not have managed with such a small difference in pressure between the inlet and the outlet.

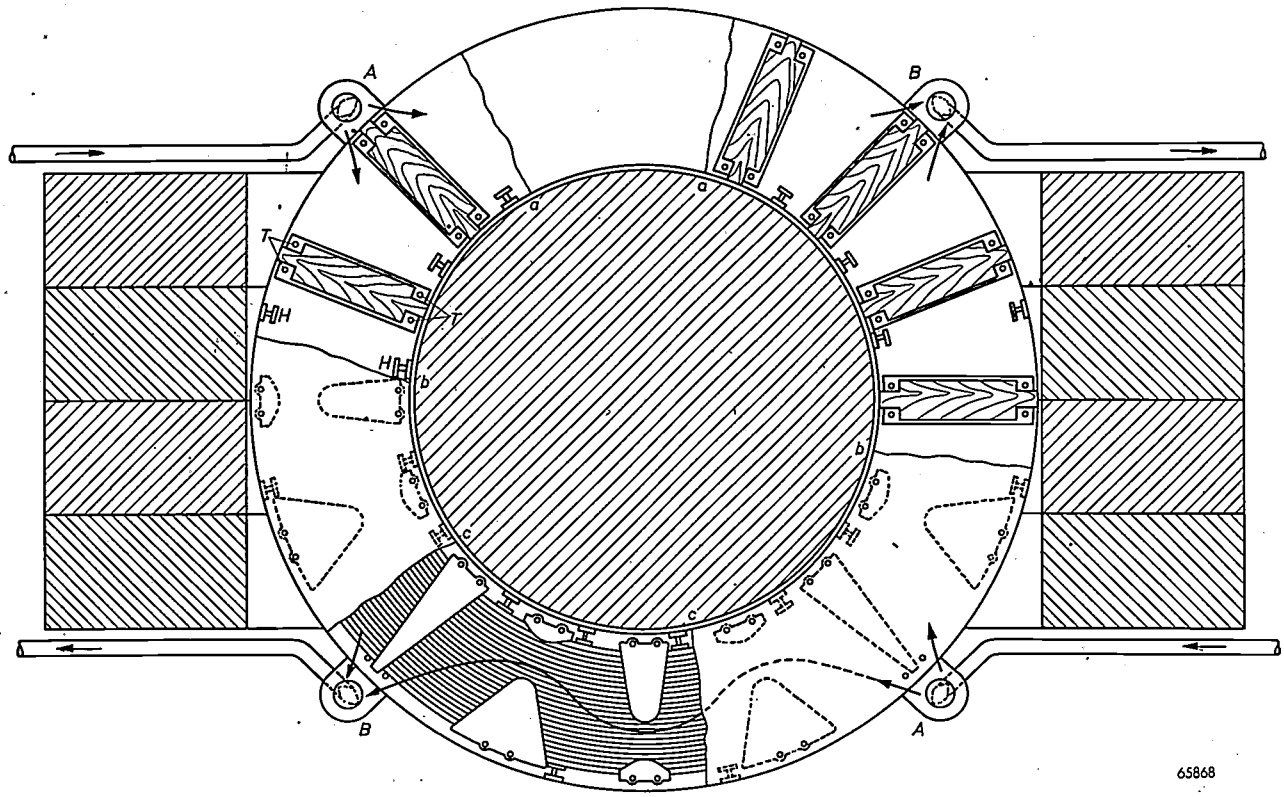


Fig. 8. Top view of one of the coils placed in position in the magnet. From *a* to *b* the cover of the coil tank has been removed to show the wooden boards taking up the pressure of the upper coil bearing against its tank cover or supporting the tank cover of the lower coil; here can also be seen the uppermost of the steel plates with which the 16 pancakes are pressed down on top of each other; *T* are the steel tie rods used for this. From *b* to *c* one looks down onto an insulating layer over the uppermost pancake, and from *c* to *c* this uppermost pancake has been cut away to show the turns of the underlying pancake and the tongue-shaped pieces of resin-bonded fabric guiding the cooling oil along a zig-zag path between the two pancakes. At *A*, *A* the oil is fed into the tank (the 15 oil channels lying one above the other between the 16 pancakes are connected in parallel), at *B*, *B* this oil is carried off, so that the oil traverses only a quarter of the circumference of a pancake. *H* are the packing posts of wood between the pancakes and the wall of the tank.

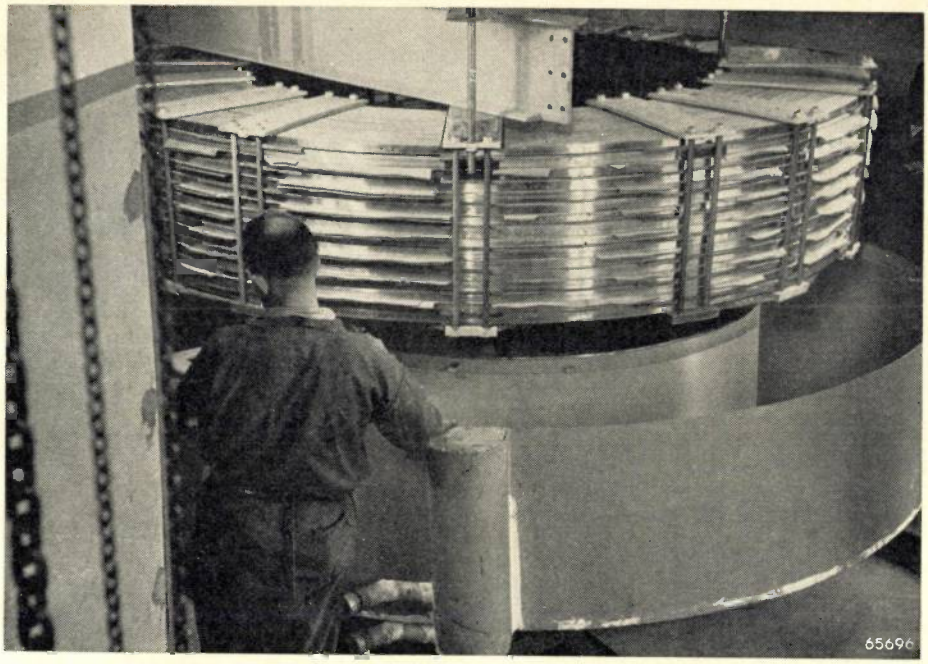


Fig. 9. A coil, consisting of 16 pancakes pressed together, ready to be placed in the annular tank.

The average temperature of the oil reaches a rather high level (about 45 °C). The equilibrium is determined, inter alia, by the decrease in viscosity of the oil with rising temperature — accompanied by an increase in the rate of flow — and by the increased heat transmission to the water with rising temperature of the oil. Since the expected temperature of the coils had already to be known for

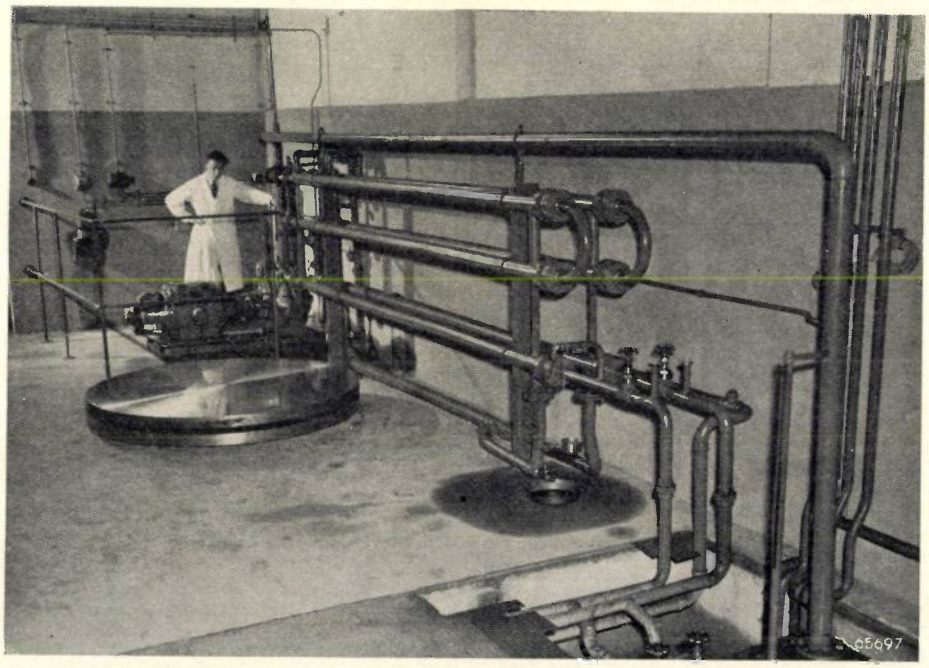


Fig. 10. Cooling plant. On the right in the foreground is the underground collecting tank for the hot oil coming from the coil tanks. The oil is pumped up (pump in the background) from the collecting tank and forced through 10 parallel copper pipes of 2.2 cm diameter and 12 m in length. These pipes are mounted, five at a time, in two steel tubes of 9 cm diameter in which cooling water flows along the oil pipes (100 l/min). These two tubes, seen in the middle of the photograph, are each folded into three horizontal lengths one above the other so as to save space. After having circulated through the cooler the oil flows back into the coil tanks.

calculating the dimensions of the magnet, this equilibrium was previously investigated with experimental models so as to ascertain, in particular, the flow resistance of the channels and the amount of heat that can be dissipated at different rates of flow of the oil.

Owing to the comparatively high temperature of the oil also the thermal radiation of the coil tanks with their large surface contributes considerably to the dissipating of the heat.

Summary. The large magnet of the Philips synchrocyclotron at Amsterdam was designed for accelerating the deuterons to an energy of 25 million electronvolts, using an air gap of 36 cm. It was known that the power for the excitation could be about 80 kW, and a steel with a certain magnetisation curve was available. The magnetic circuit — the general shape of which is prescribed by various technical considerations — had to be so dimensioned as to keep the total cost of steel and copper to the minimum. These optimum dimensions cannot be deduced exactly but have to be found approximately by experimental calculations and measurements taken with models. The most important considerations determining the design procedure are explained in this article. With the dimensions finally chosen — a pole diameter of 1.80 m, a yoke length of 4.70 m and a yoke height of 3.50 m — 497,000 ampere-turns are needed to obtain on the axis of the air gap a flux density of 1.35 Wb/m²; 76% of these ampere-turns are required for overcoming the reluctance of the air gap, the remainder being required for the reluctance of the steel circuit. The total weight of steel used for the magnet is 200 tons and that of copper 30 tons. The magnet has been built up from 16 steel blocks and 4 steel pole pieces, which, with the exception of the upper pole suspended from bolts, have been joined together without any fixing means whatever. The coil, made in two parts, has been calculated for a total voltage of about 500 volts (resistance about 3 ohms, total number of turns over 3000). Each part is built up from 16 "pancakes" of wound copper strip with paper insulation. The pancakes are cooled by circulating oil; the hot oil is pumped into a cooler where the heat is carried off by a stream of water (water consumption 100 litres/min). With the cooling installation constructed it was found possible to use a slightly stronger excitation current for the magnet (coil temperature about 45 °C), so that now the apparatus can be normally worked with a flux density of 1.38 instead of the planned 1.35 Wb/m². In this article the measures are also described by means of which a radial variation of the flux density is reached which is suitable for stabilizing the orbits described by the particles. The largest useful orbit appears to have a radius of 78 cm. Owing to this and the slightly greater flux density, deuterons of 28 MeV can be produced with this synchrocyclotron instead of the planned 25 MeV.

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

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

1953: J. M. Stevels: Le fer dans le verre (Verres et Réfractaires 4, 293-298, 1950, No. 5). (Iron in glass; in French.)

The behaviour of iron in glass is intimately related to the variable valence of the iron atom, which makes electronic transitions possible. In network-forming positions Fe^{3+}-ions cause a brownish colouring, while in network-modifier positions there is practically no colouring. Fe^{2+}-ions seldom occur alone and, if so, mostly as network modifiers, but sometimes also as network formers. These ions do not cause a visible colouring. The bluish-green colour of some glasses containing iron is due to the simultaneous action of Fe^{2+}- and Fe^{3+}-ions.

Indications exist that at least one of these ions (or both) must be present in network-forming positions. The influence of Ti^{4+} is of importance in this respect because this ion shifts the equilibrium between the network-former and the network-modifier positions of Fe-ions in the latter direction.

It may be that decolouring of glass simply means elimination of TiO_2. A further study of the equilibrium $Fe^{2+} \rightleftarrows Fe^{3+}$ in both positions is needed.

1954*: A. Claassen: Titan (Hb. Anal. Chem., Dritter Teil, Band IVb, Springer, Berlin 1950; pp. 1-169).

Report on quantitative analytical methods for the determination of Ti and the various methods of separation.

1955*: A. Claassen: Zirkon und Hafnium (Hb. Anal. Chem., Dritter Teil, Band IVb, Springer, Berlin 1950, pp. 170-287).

Report on quantitative methods for the determination of Zr and Hf and the current methods of separation.

1956: J. L. Snoek: Density variations in aluminium (Phil. Mag. **41**, 1188-1192, 1950, Nov.)

From careful determinations of the density of samples of aluminium in various conditions the conclusion is drawn that grain boundaries in aluminium may have a more spacious structure than is generally assumed. In an addendum the possible influence of gas bubbles (H_2) on the density is discussed.