

DIELECTRIC LOSSES IN GLASS

by J. M. STEVELS.

666.1:537.226.3

One of the many industries in which glass is used is that of electrical engineering. A problem arising in this industry is the behaviour of the glass in an electric alternating field. The dielectric losses of glass form one of the factors that decide this behaviour. Extensive investigations are still being carried out with a view to producing glasses having small dielectric losses at all frequencies. These investigations have not only led to an improvement of the existing kinds of glass but they have given a better insight into the physical and chemical structure of glass.

Dielectric losses

Glass is a material that is much used in electrical engineering: the bulbs of incandescent lamps and of radio valves (transmitting and receiving valves) are made from glass, that is also indispensable as an insulating material and for very many other purposes. It is therefore of great importance to know how the electrical properties of glass are affected by its composition, for this enables a choice to be made of the best kinds of glass for certain uses. This article will deal with the behaviour of glass in an electric alternating field, attention being mainly focused on the dielectric losses of glass.

When a piece of glass is interposed as a dielectric between the plates of a capacitor and an alternating voltage with a frequency f is applied to that capacitor, if the glass were free of losses, then the current flowing through the circuit would be exactly an angle $\pi/2$ in advance of the voltage across the capacitor. This means that no energy would be dissipated in the capacitor. In practice, however, every dielectric, and thus also glass, dissipates a smaller or larger proportion of the energy applied and converts it into heat. The measure of this dissipation of energy (the dielectric losses) is denoted by the angle δ by which the actual difference in phase between current and voltage deviates from $\pi/2$. The energy dissipation per unit of time for a small δ is approximately equal to $f \tan \delta$; the angle δ is called the loss angle.

In glass, as is the case with other materials, the dielectric losses vary strongly with the frequency of the electric field. The "spectrum" of the losses of glass is broadly represented by the fully drawn curves in *fig. 1*, which have been drawn for two temperatures, viz. 50 °K and room temperature. It is to be noted, however, that these diagrams have no quantitative significance; the losses differ greatly according to the kind of glass considered. The graph gives only a general impression of the losses as a function of frequency. Although the frequency

range in these diagrams extends from 1 to 10^{14} c/s, only measurements between the frequencies of 50 c/s and 10^{10} c/s will be dealt with.

Fig. 1 reveals that there are two maxima of the losses, shifting in frequency as the temperature is changed but not in the same direction. Between the maxima is a "valley" of relatively small losses, the depth of this valley decreasing as the temperature drops.

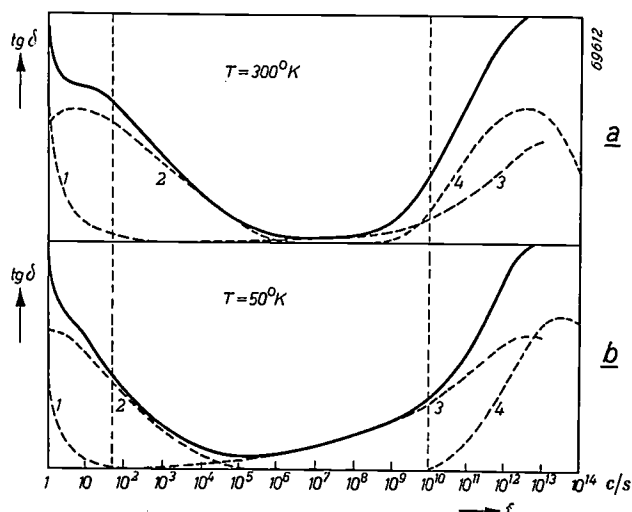


Fig. 1. Diagrammatic representation of the frequency spectrum of the dielectric losses in glass; *a* gives the spectrum at approximately room temperature, *b* shows how the spectrum appears at a very low temperature. The fully-drawn curve gives the total losses. It can be analyzed into four kinds of losses occurring at different frequencies, as represented by the broken-line curves: 1 conduction losses, 2 relaxation losses, 3 deformation losses, 4 vibration losses. It is seen how the vibration losses, occurring at high frequencies, are displaced to still higher frequencies as the temperature falls. The conduction and relaxation losses, on the other hand, are displaced towards lower frequencies, while the deformation losses increase over the whole line. These diagrams are of only theoretical significance. The frequency range here extends from 1 to 10^{14} c/s; this article deals only with the range between 50 and 10^{10} c/s (vertical broken lines).

From what is already known about the behaviour of the dielectric losses in other solids an attempt may be made to analyze the loss spectrum of glass

into various components, taking into account the specific property distinguishing glass from crystalline materials, namely the relatively great disorder in the orientation of the ions from which glass is built up.

It will be seen that there are differences in the manner in which the various kinds of losses depend upon the composition of the glass. Thus from the analysis of the losses some idea can be obtained as to how the composition has to be varied in order to reduce or increase the losses at certain frequencies.

The result of the analysis has been graphically represented in fig. 1 by the broken-line curves. Four kinds of losses are to be distinguished, named respectively conduction losses, relaxation losses, deformation losses and vibration losses, which will now be dealt with in that order.

Conduction losses

Glass is not a perfect insulator. Under the influence of an electric force the network-modifying ions move through the whole of the network ¹⁾, and in doing so they give off part of the energy obtained from this force to the network in the form of heat.

The conduction losses (curve 1 in fig. 1) depend upon the conductivity σ according to the formula

$$\tan \delta = \frac{\sigma}{\omega \epsilon' \epsilon_0}, \dots \dots \dots (1)$$

where $\omega (= 2\pi f)$ represents the angular frequency of the electric alternating field and ϵ' the dielectric constant of the glass (ϵ_0 is the dielectric constant of vacuum). Thus $\tan \delta$ is inversely proportional to the frequency (since δ and ϵ' are not frequency dependent, or scarcely so), and the energy dissipation per time unit does not depend upon the frequency. This is a well-known property of a conductor. When $\tan \delta$ for a lime glass ($\sigma = 10^{-10} \Omega^{-1} \text{m}^{-1}$, $\epsilon' = 6$) is calculated according to the above formula for a frequency of 1000 c/s one finds the small value of 3×10^{-4} , whereas an actual measurement gives $\tan \delta = 100 \times 10^{-4}$, so that at this frequency there must be other causes of losses much more important than conduction. As a general rule, for frequencies higher than 50 c/s the conduction losses are negligible compared with the other losses. In this article, therefore, conduction losses will no longer be considered.

¹⁾ The structure of glass, and especially the conceptions of network-modifying ions and network-forming ions, have been dealt with in a previous article recently published in this journal (J. M. Stevels, Philips Techn. Rev. 13, 293-300, 1952, No. 9), which will further be referred to as article I.

Relaxation losses

It is not only the transport of network modifiers through the whole piece of glass considered that causes losses, but also the transport over atomic distances in the glass plays a part. In fig. 2 the variation of the potential energy of the ions in an arbitrary part of the network is indicated for an arbitrary direction (x) in that network.

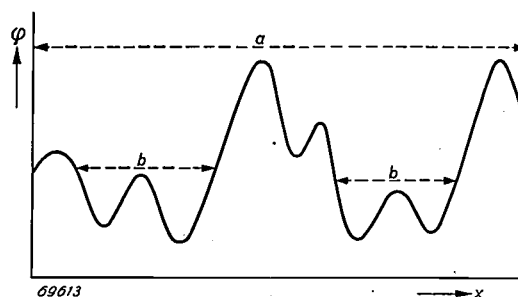


Fig. 2. The undulated curve represents the potential ϕ of the network-modifying ions along an arbitrary direction in the glass. The movement indicated by a can only be carried out by high-energy ions and corresponds to conduction, while the movement b is carried out by ions with less energy and results in relaxation losses.

The successive potential barriers, with varying heights and distances, form obstacles preventing an ion from travelling further through the glass. The potential barriers arise partly from electrical causes (mutual repulsion of the ions) and partly from purely geometrical causes: the movement of the ion concerned may be obstructed by the other ions. Similar potential barriers are found also in crystals, but there the distances and heights are very regularly distributed.

When an electric field is applied the ions tend to move in the direction of this field. Usually the field alone will not be sufficient to carry the ions over the potential barriers. For that to take place the ions have to take up extra energy, and this is obtained from collisions with other ions due to the thermal movement in the glass. It takes some time, however, before the ions have accumulated sufficient energy to pass over a potential barrier; this time is called the relaxation time τ . If the electric field is an alternating field with angular frequency ω , and ω is not large compared with $1/\tau$, then during each half cycle the ions will have ample time to jump over one or more potential barriers, thereby gaining energy from the field, which is transmitted to the network in the form of heat; one then speaks of losses. At very low frequencies ($\omega \ll 1/\tau$) these losses are small, because then the ions already jump when the instantaneous value of the field differs but little from zero: the energy gain of the ions is then small and the lattice receives little energy.

When, however, ω becomes of the order of $1/\tau$ then, on the average, the ions will not pass over the potential barrier until the field has practically reached the maximum value. The ions then gain much energy and the losses are large. If the frequency of the alternating field is much higher than $1/\tau$ the ions jump independently of the phase of the alternating field and the losses are again small.

The relation between these relaxation losses and the angular frequency ω of the field is expressed as

$$\tan \delta = \frac{\omega\tau}{1 + \omega^2\tau^2}, \dots \dots \dots (2)$$

which is entirely in accordance with the foregoing reasoning: the losses occur only at angular frequencies round about $1/\tau$.

As may be understood from fig. 2, in glass there is more than one relaxation time τ , since the local structures differ considerably. This makes the relaxation losses perceptible over a relatively rather wide range of frequencies. The spectrum of these relaxation losses (broken-line curve 2 in fig. 1) shows a wide peak, the limits of the range in which the losses occur lying, at room temperature, at about 10^{-3} and 10^6 c/s. As already remarked, at frequencies higher than 50 c/s the relaxation losses exceed the conduction losses considerably.

The conduction losses could also be regarded as a kind of relaxation losses. The ions which collect sufficient energy to pass over the highest potential barriers will be able to travel through the whole piece of glass. In order to pass over the highest barriers, a very long relaxation time will obviously be required, and in that case formula (2) becomes

$$\tan \delta \sim \frac{1}{\omega\tau}, \dots \dots \dots (3)$$

$\tan \delta$ being inversely proportional to the frequency, a relation applying, indeed, for conduction losses.

Deformation losses

Relaxation losses are therefore due to an after-effect phenomenon. There is reason to assume that another kind of losses occur in glass which bear the character of after-effect losses. These are denoted by the name of deformation losses to distinguish them from the relaxation losses just dealt with. Deformation losses do not arise from the movement of individual network modifiers, but from the movement of whole sections of the network, in particular a sort of kinking movement of the chains of the network (see I). As is rather obvious, such a phenomenon has a very short relaxation time: it is to be imagined that the chains make movements

resembling the jumping of a bent leaf spring. Here again there are a large number of different relaxation times τ . Since on the average τ is very small, for the range of not too high frequencies formula (2) may be written approximately (for the case where there would be only one relaxation time) as:

$$\tan \delta \sim \omega\tau. \dots \dots \dots (4)$$

From this formula it follows that the deformation losses increase about proportionately with the frequency and in most cases will be very small. The spectrum of the deformation losses is represented in fig. 1 by the broken-line curve 3.

Vibration losses

The fourth kind of dielectric losses to be expected in glass are due to a resonance phenomenon. The ions in the glass — both the network formers and the network modifiers, as well as the oxygen ions — may vibrate with a certain frequency round about their state of equilibrium, and this they will in fact do as a consequence of the thermal movement. Regarding the ion as a harmonic oscillator, its frequency of vibration may be approximately represented by the formula

$$\omega_{res} = \sqrt{\frac{a}{M}}, \dots \dots \dots (5)$$

where a is the proportionality factor denoting the relation between the restoring force and the displacement from the state of equilibrium, and M is the mass of the ion in question. Consequently ions of different mass and ions at different places in the crystal (with different a) will usually vibrate with different frequencies.

When an electric force is applied with a frequency approximately equal to the vibration frequency of an ion, then resonance may occur. Since the vibrations of the ions are always damped, this resonance is accompanied by losses. If all the ions had the same resonant frequency, then the spectrum of these vibration losses would show a maximum round about that frequency. But, as already observed, there are a large number of resonance frequencies. This, and the possibility of the vibrations of the ions being strongly damped, leads to a wide maximum in the spectrum of the vibration losses. This spectrum is represented in fig. 1 by the broken-line curve 4. Unfortunately, with one single exception, the actual maximum has never yet been determined, since measurements so far carried out have not gone beyond 10^{10} c/s.

Vibration losses in glass are directly comparable to the well-known infra-red absorption in polar crystals, which arises from a similar resonance phenomenon. In the case of a crystal like rock salt this absorption takes place at about 10^{13} to 10^{14} c/s, where the maximum is very narrow. The fact that in the case of glass "spurs" of the vibration losses are to be observed at much lower frequencies (about 10^{10} c/s) may be ascribed, as already explained above, on the one hand to the presence of ions with low resonance frequencies and on the other hand to the possibility of a strong damping of the ion vibrations in glass, which leads to a considerable widening of the loss spectrum. As may be understood theoretically, strongly damped resonance vibrations form, as it were, a transition to the relaxation vibrations previously discussed. The question in how far the vibration losses in glass are to be regarded purely as a resonance phenomenon or as a sort of after-effect phenomenon can to a certain extent be answered, as will presently be shown, when the temperature dependency of the losses is investigated. It is to be added, however, that in order to gain a deeper insight into these phenomena it is necessary to study the losses in glass in the infra-red at the frequencies of 10^{11} to 10^{13} c/s unfortunately it is extremely difficult to carry out absorption measurements in that range.

Having now dissected the total loss spectrum in fig. 1 into four components, the spectrum can be well understood qualitatively. For a quantitative investigation great difficulties have to be overcome, due to the irregular structure of glass. Some confirmation of the picture formed is, however, to be obtained when we come to investigate the temperature dependency of the various kinds of losses. It is then found that their behaviour is roughly as represented in figs 1a and b, the relaxation and conduction losses being displaced towards lower frequencies as the temperature drops, while the deformation losses increase over the whole line. As to the vibration losses, there are indications that with falling temperature these are displaced towards higher frequencies, and this would be in agreement with what is to be theoretically expected for this kind of loss.

Influence of temperature upon the losses

The foregoing conclusions may be briefly explained as follows.

The temperature dependency of conduction losses is governed entirely by that of the conductivity. As is presumably known, for a fixed frequency the conductivity σ decreases with falling temperature, and thus, according to formula (1), also the conduction losses decrease.

The relaxation times τ occurring with relaxation losses and deformation losses will be greater as the temperature T falls, since the mutual collisions of the ions are then fewer and weaker and it takes longer for the ions to accumulate the energy required to overcome a potential barrier. An increase in τ means a displacement of the losses to lower frequencies, as is indeed observed in the case of relaxation losses when T drops.

As to the deformation losses described by formula (4), an increase in τ means that the losses increase over the whole of

the frequency range with which we are concerned. For a quantitative calculation of the temperature dependency of relaxation and deformation losses it is necessary to take into account the great variety of relaxation times occurring in glass, but in their generality the conclusions still hold. To illustrate the relation between the losses at medium frequencies and the temperature, in fig. 3 a curve has been plotted representing $\tan \delta$ of a lime glass as a function of temperature at the frequency 1.5×10^6 c/s. It appears that, for temperatures round about 70 °K, $\tan \delta$ shows a broad maximum, then becoming rather small at about 150 °K and rising again rapidly at temperatures higher than 200 °K. The broad maximum is due to the deformation losses, while the rise above 200 °K is to be ascribed to relaxation losses, as may be understood from the following.

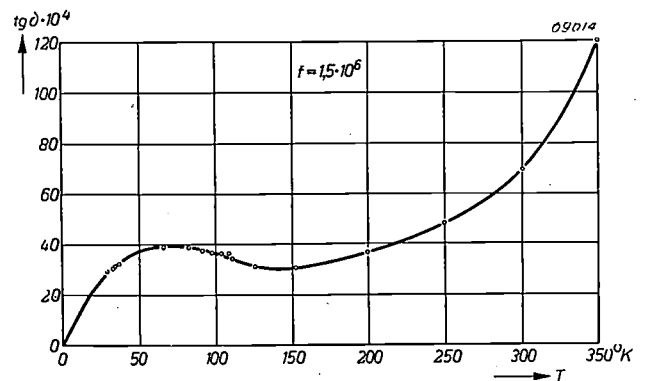


Fig. 3. Dielectric losses in a lime glass at the frequency 1.5×10^6 c/s as a function of temperature. The maximum at low temperatures corresponds to the deformation losses, the rise of the curve on the right is due to relaxation losses.

For a fixed frequency it is possible to determine the temperature at which the deformation losses have reached their maximum. The relaxation times are always much smaller for the deformation losses than for the relaxation losses, so that the frequency mentioned will be situated on the right-hand descending part of the frequency curve for the relaxation losses. When the temperature rises both the spectrum of the deformation losses and that of the relaxation losses will be displaced towards higher frequencies. The maximum of the deformation losses will, therefore, move away from the fixed frequency, whereas the maximum of the relaxation losses moves towards this frequency.

A note is to be added here. The conduction loss is also a kind of relaxation loss, with such a great relaxation time as to occur always on the right-hand descending part of the curve representing the relation between losses and frequency. The conclusion already arrived at in another way, that these conduction losses diminish with falling temperature, is therefore not surprising when considered also from this point of view.

Purely vibrational losses will be displaced towards higher frequencies as the temperature falls. This is because the vibrating ions have on the average higher resonant frequencies at lower temperatures, as may be understood from quantum theory. It has already been pointed out that the vibration losses in glass probably form a transition to relaxation losses. Since the temperature dependency of relaxation losses is just the reverse of that for vibration losses, it is not possible to predict the influence of temperature on these losses at the highest frequencies without carrying out further experiments

in the far infra-red. The indication, already mentioned, that with falling temperature these losses are displaced to higher frequencies makes it highly probable, however, that the character they bear is preponderantly that of resonance.

We now have to consider the influence of the composition of the glass upon the different kinds of losses. First its influence upon the relaxation losses will be dealt with, and then the influence upon the vibration losses; as mentioned above, the conduction losses will be disregarded, since, whatever the composition of the glass may be, for frequencies above 50 c/s they are of no importance; as to the deformation losses too little is yet known about them.

Influence of the chemical composition of glass upon relaxation losses

The manner in which relaxation losses are affected will be considered under five headings.

Nature of the ions

As a rule only those network modifiers which readily migrate through the glass will be capable of yielding a considerable component towards the relaxation losses. The mobility of the ions in a given potential field is determined mainly by their size: it is particularly the relatively small Li^+ and Na^+ ions which play an important part in this respect. Glasses containing a large proportion of these ions show high relaxation losses. The K^+ ion is so much larger that it contributes much less towards the losses. By way of illustration mention may be made of a series of low-melting glasses of the composition 53.3 mol % SiO_2 , 10.2 mol % PbO , 4.5 mol % CaF_2 and 32 mol % M_2O , with M representing an alkali metal. For a frequency of 1.5×10^6 c/s and at a temperature of 20 °C, the value of $\tan \delta$ for $\text{M} = \text{Li}$ is 132×10^{-4} , for $\text{M} = \text{Na}$ $\tan \delta$ is 106×10^{-4} and for $\text{M} = \text{K}$ it is 54×10^{-4} .

Divalent ions, owing to their larger charge, are bound more strongly to their surroundings and thus are usually less mobile. An exception is the small Mg^{2+} ion which appears to be capable of yielding a considerable contribution towards the losses.

From now on the Li^+ , Na^+ and Mg^{2+} ions will be denoted by the name of mobile ions, whilst the divalent network modifiers, such as Ca^{2+} , Ba^{2+} , Pb^{2+} , will be called immobile ions.

It will be clear that the losses depend also largely on the nature of the network. This influence can be reduced to three effects.

Influence of R

As explained in article I, the ratio R of the number of oxygen ions to the number of network-forming ions is a measure for the coherence of the network from which the glass is built up. The larger the ratio R , the less coherent is the network and the less dense is the structure of the glass. This lends more mobility to the network modifiers, and the relaxation losses may be expected to increase.

Influence of the addition of immobile ions

If metal ions showing little mobility — either because they are large or because they are strongly bound electrically — are added to a glass of a certain composition, then there will be less possibility for displacement of the mobile ions. Thus, given equal concentration of mobile ions, the addition of immobile ions to a glass reduces its loss angle.

The effects mentioned here can be quantitatively described by the formula :

$$\tan \delta = A C_m R (1 - BC_i), \dots (6)$$

where C_m represents the concentration of the

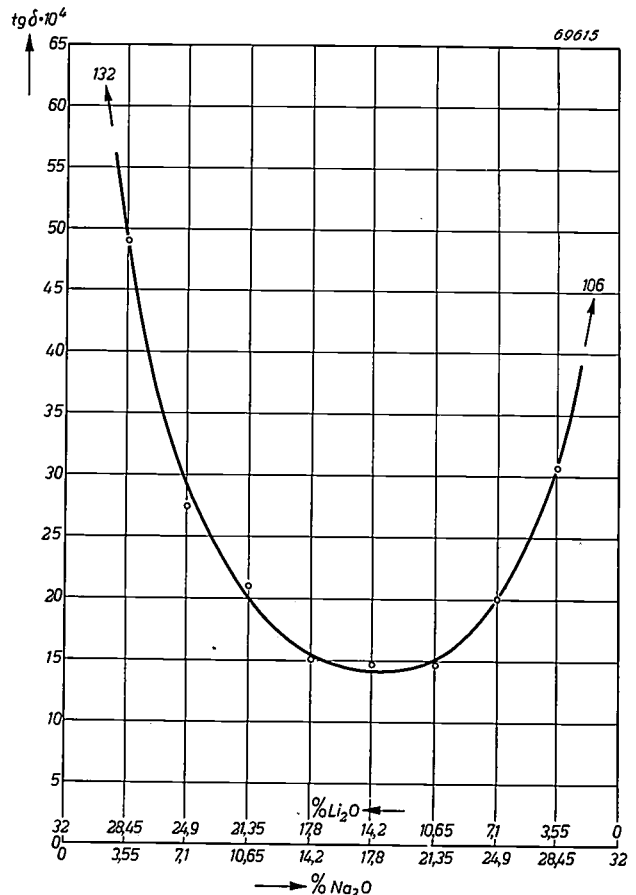


Fig. 4. Replacement of part of the Li^+ ions in a lithium-silicate glass by Na^+ ions gives a very rigid structure and thus reduces the relaxation losses. If, however, this replacement is carried too far the structure becomes less rigid again. The curve for $\tan \delta$ therefore shows a minimum at about equal concentrations of Na^+ and Li^+ ions (measured at the frequency of 1.5×10^6 c/s).

mobile ions and C_i that of the immobile ones; the coefficient A depends on the frequency and temperature; B is a geometrical factor²⁾).

Difference in packing, with equal R

If in a glass containing only one kind of network modifiers, some of these ions are replaced by other network modifiers having a different radius, then a more compact network can be obtained. As the glass cools from the melt, the mutually coherent chains of network-forming ions will then preferentially so arrange themselves that the largest interstices in the network are occupied by large network modifiers and the smallest interstices by small network modifiers. The resulting denser structure leads to a reduction of the losses. This is illustrated in *fig. 4*, relating to the glass of the composition mentioned in the beginning of this section, with M representing partly lithium and partly sodium. The value of $\tan \delta$ (measured for a frequency of 1.5×10^6 c/s) varies with the variation of the proportions of Na^+ and Li^+ ; the losses are

the smallest with about equal molecular concentrations of Na^+ ions and Li^+ ions.

Relaxation losses in borate glasses

In article I it has already been discussed how in certain respects the structure of borate glasses differs from that of other glasses. In a glass of the composition $x \text{Na}_2\text{O} \cdot y \text{B}_2\text{O}_3$, if $x < 18$ mol % (the

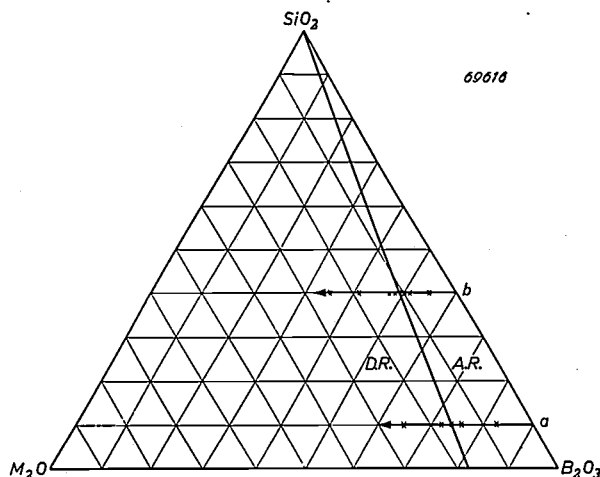


Fig. 5. Phase diagram of the ternary system $\text{M}_2\text{O} \cdot \text{B}_2\text{O}_3 \cdot \text{SiO}_2$, the borosilicate glasses. M represents an alkali metal. Two regions can be distinguished: the accumulation region $A.R.$ and the destruction region $D.R.$

²⁾ See, e.g., J. M. Stevels, Progress in the theory of the physical properties of glass, Elsevier Publ. Co., Amsterdam 1948, p. 76.

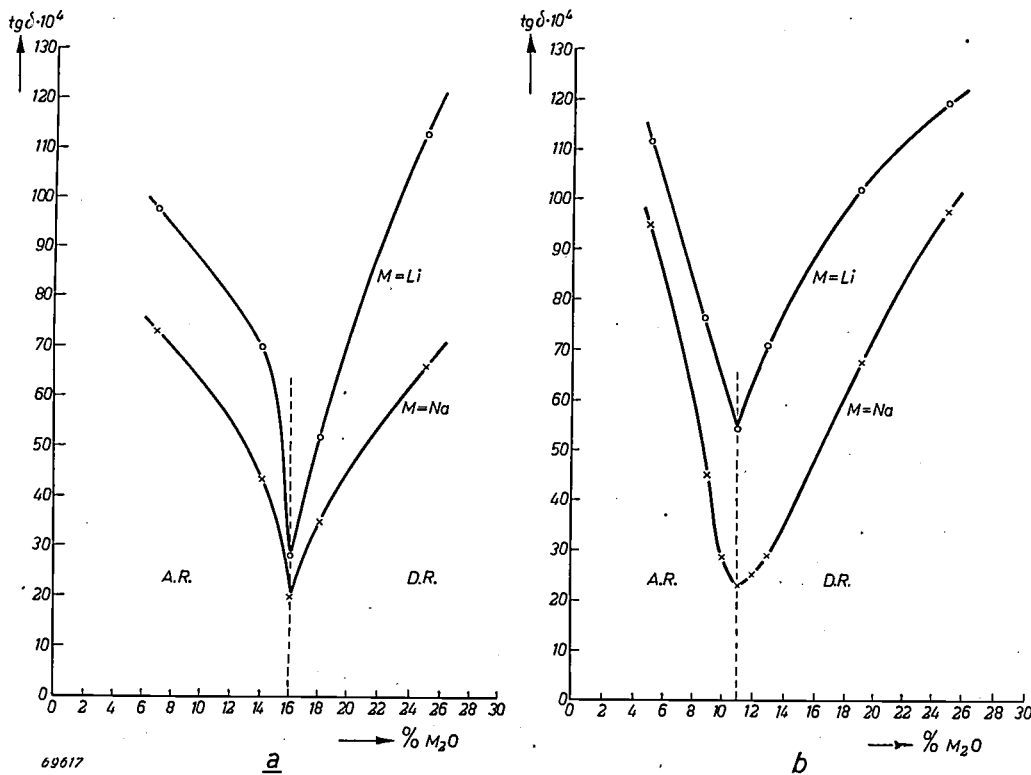


Fig. 6. Variation of $\tan \delta$ with the Na_2O or Li_2O content of borosilicate glasses; *a*) corresponds to the compositions *a* in *fig. 5*, *b*) corresponds to the compositions *b*. The curve for $\tan \delta$ shows a sharp minimum on the $A.R.$ - $D.R.$ separating line. It is also seen that Li ions cause a larger loss angle than Na -ions, because the former are more mobile. Measurements were taken at the frequency of 1.5×10^6 c/s.

accumulation region), the network-modifying Na^+ ions are incorporated in the network without any breaking of oxygen bridges, and thus there are no non-bridging oxygen ions. This is due to the fact that the B^{3+} ions may occur both in the centre of an oxygen triangle and in that of an oxygen tetrahedron. With larger concentrations of Na_2O , however, oxygen bridges tend to be broken, and this range of concentrations is called the destruction region.

With increasing concentration of Na_2O in the accumulation region, the network becomes more and more rigid until the maximum rigidity is reached just at the transition concentration of 18 mol % Na_2O , above which it begins to lose its rigidity again. Such has been illustrated in article I with reference to the behaviour of the expansion coefficient of the glass as a function of its composition.

A similar phenomenon arises in the case of borosilicate glasses. In *fig. 5* a phase diagram is given for the system $\text{M}_2\text{O}-\text{B}_2\text{O}_3-\text{SiO}_2$, where M stands for a monovalent network modifier. This diagram can be divided into an accumulation region *A.R.* and a destruction region *D.R.* The compositions found along the line separating these two regions are characterized by a maximum rigidity of the structure.

Obviously the dielectric losses (conduction and relaxation losses) will be influenced by this change in structure. As was to be expected, the losses show a decided minimum just on the dividing line, as is illustrated in *fig. 6*. *Fig. 6a* corresponds to the compositions indicated by the arrow points *a* in *fig. 5*. *Fig. 6b* applies for the compositions *b* in *fig. 5*. Both Li^+ and Na^+ have been chosen as monovalent network-modifying ions, thus giving two perfectly analogous curves.

Influence of the chemical composition of the glass upon the vibration losses

It is possible to shift the frequency range of the vibration losses by changing the composition of the glass. The influence of the chemical composition will again be considered from a number of points of view.

Nature of the ions

One of the most important factors affecting the frequency of the vibration losses is the mass of the ion (cf. formula (5)). The heaviest ions will give the lowest resonant frequency; thus the addition of heavy ions causes the vibration losses to be displaced to lower frequencies. The charge of the ions plays a less important part (this occurs in the

restoring force constant α , which expresses the interaction of the ion with the surroundings).

Influence of R

The larger the ratio R , the less dense is the network. The ions become more loosely bound to their surroundings and the resonant frequencies are

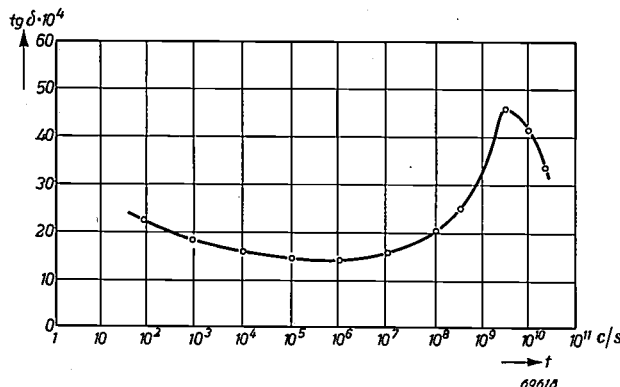


Fig. 7. Example of a glass (phosphate glass) in which the resonant frequencies lie very low, within the frequency range accessible for measurements.

shifted to lower frequencies. Phosphate glasses are of a kind in which many non-bridging oxygen ions occur and thus the ratio R is high. *Fig. 7* shows that in a certain phosphate glass the resonant frequencies are shifted so far that the maximum of the

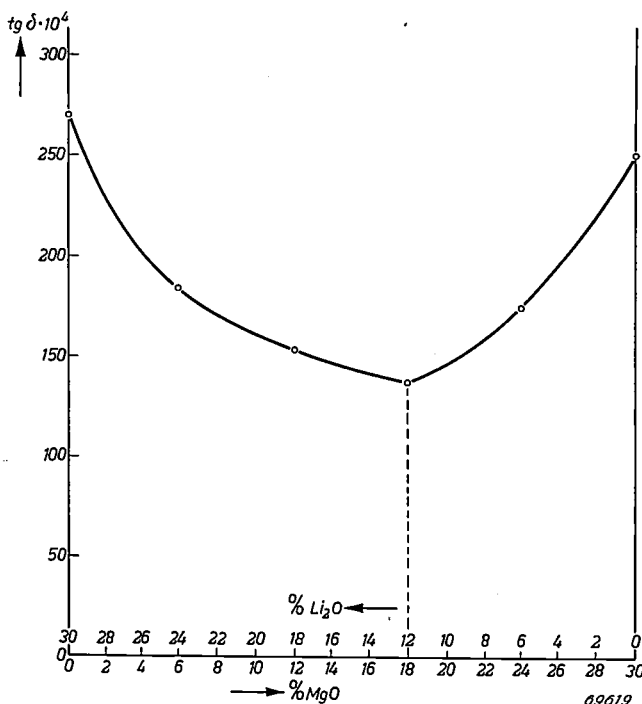


Fig. 8. The more compact structure of a silicate glass due to the presence of different kinds of network modifiers leads to a reduction of the vibration losses, measured at a frequency of 2.4×10^{10} c/s. This diagram is fully comparable with *fig. 4*. It is to be noted that the value of R in this diagram, going to the right, drops somewhat, the reason for this being that Mg^{2+} ions, if present in large concentrations, have the tendency to act as network formers. This diagram relates to a glass of the composition 70% SiO_2 and 30% ($\text{Li}_2\text{O} + \text{MgO}$).

vibration losses is situated below 10^{10} c/s (this is so far the only case where it has been possible to measure such a maximum).

The addition of some metal oxides, such as Al_2O_3

in the neighborhood of the transition the network is very rigid, and thus the resonant frequencies are very high. The "spur" of the vibration zone that can be measured is then small. This is illustrated in fig. 9.

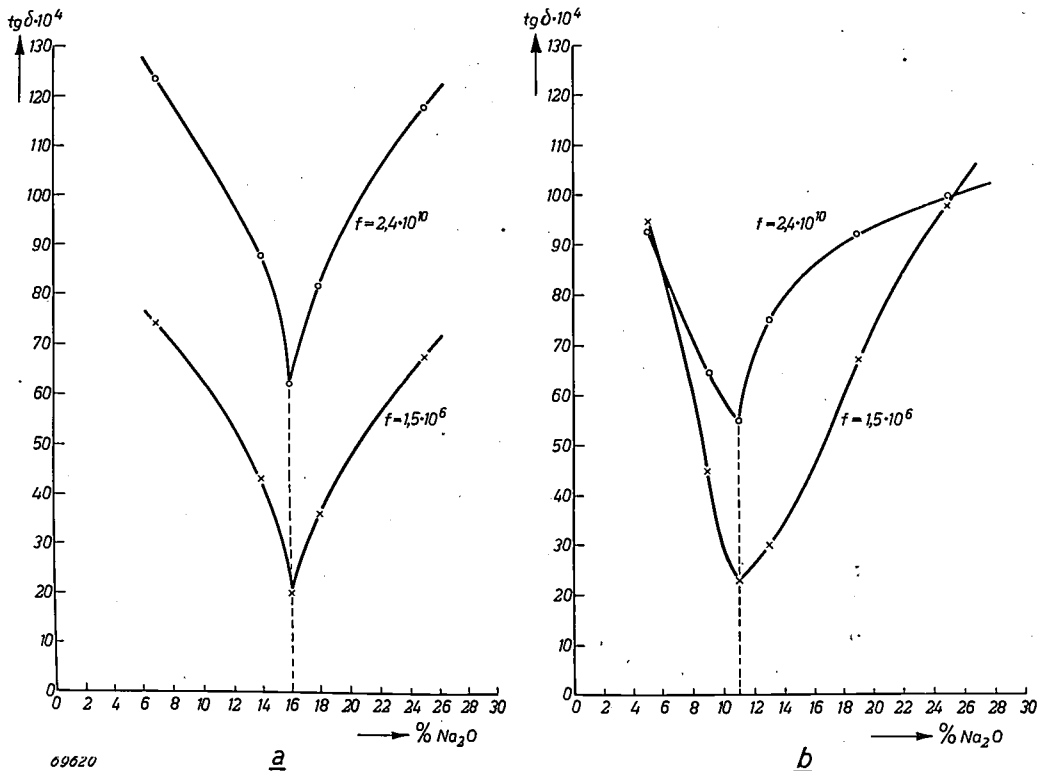


Fig. 9. The vibration losses in borate glass, just as the relaxation losses, are affected by changes in composition going from the A.R. to D.R. region. Fig. 9a relates to the compositions a in fig. 5, fig. 9b to the compositions b, with Na^+ as network modifier. The vibration losses have been measured at 2.4×10^{10} c/s; for comparison also the relaxation losses measured at 1.5×10^6 c/s are reproduced from fig. 6.

and ZnO , which reduce R , will lower the losses at higher frequencies, as follows from the foregoing.

Difference in packing, with equal R

As already seen when discussing the relaxation losses, the structure of the network becomes more compact the more the network-modifying ions occurring in the glass differ from each other in their dimensions, even when R remains constant. The more compact the network, the higher are the resonant frequencies of the vibration losses and thus the less are these losses noticed in the range of the relatively low frequencies where the measurements are taken. An illustration of this is given in fig. 8, which is directly comparable with fig. 4.

Vibration losses in borate glasses

The effect of the transition from A.R. to D.R. upon the vibration losses is exactly comparable to the effect upon the relaxation losses. In the neigh-

Some examples of the spectrum of dielectric losses in glass

Fig. 10 gives the $\tan \delta$ as a function of frequency for a number of glasses. The region of the relaxation (and conduction) losses on the left and that of the "spurs" of the vibration losses on the right are clearly distinguished. The glass A contains a large proportion of mobile ions and thus shows very large losses. The fact that the curve D, for lead glass, rises relatively more steeply at high frequencies than the other curves can be explained, according to the foregoing, by the presence of a large number of heavy ions, as a consequence of which the resonant frequencies lie lower than those in the other glasses.

Fig. 11 shows the losses of a lead-containing glass at a number of frequencies as a function of temperature. This is to be compared with fig. 3. The influence of the deformation loss at low temperatures is clearly seen. This influence is of

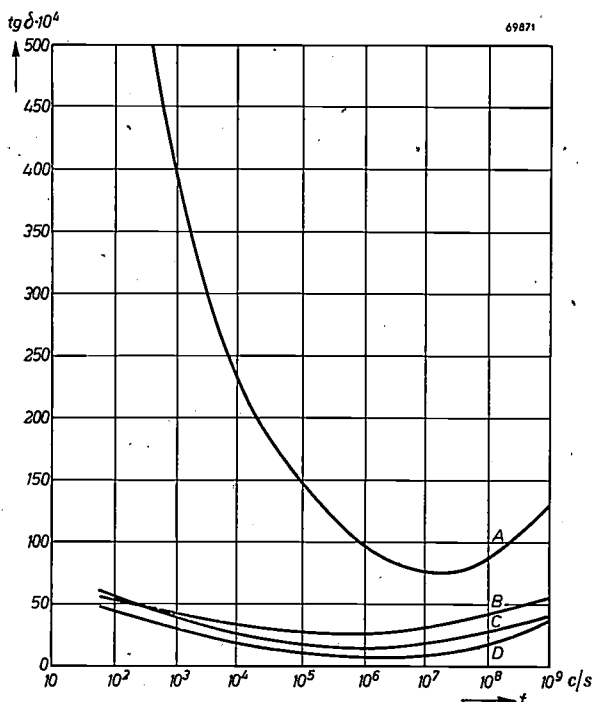


Fig. 10. Loss spectra of a number of glasses. *A* soda-lime glass; the losses, especially the relaxation losses on the left, are very high, owing to the large proportion of mobile Na^+ ions contained in the glass. *B* alkali-free silicate glass, and *C* borosilicate glass: these glasses have low losses. *D* lead glass: the losses for this glass show a relatively steep rise at higher frequencies, because the resonant frequencies in this glass are low.

particular importance for higher frequencies, as follows also from the theory. The increase in the losses at higher temperatures is again due to relaxation losses.

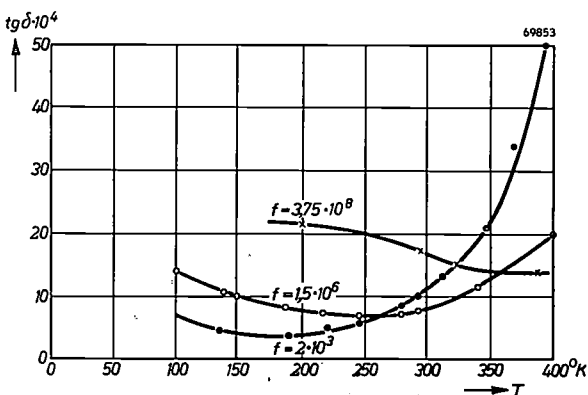


Fig. 11. Loss angle of the glass with the composition 45% SiO_2 , 32% PbO , 5% CaF_2 , 4% Na_2O and 14% K_2O , as a function of temperature, at different frequencies. The deformation losses (on the left of the diagram) are relatively considerable (cf. fig. 3), particularly at high frequencies.

Practical applications

Glasses with low dielectric losses

Although in practice glasses with high dielectric losses are sometimes desired for special purposes, these will not be discussed here. It has already been made sufficiently clear how such glasses can be produced.

Of most importance are the glasses which show small dielectric losses at normal working temperatures; these are used for the construction of all sorts of electronic tubes and valves. From the foregoing comments it follows that a distinction has to be made between (1) glasses intended for use at intermediate frequencies (radio valves), where the dielectric losses are mainly relaxation losses, and (2) glasses intended for constructions where very high frequencies are applied and mainly vibration losses occur. These are two entirely different fields; a glass that is satisfactory in one field need not at all be so in the other. It is certainly not difficult to find a glass with low dielectric losses at all frequencies — such as fused silica — but the difficulty lies in the fact that for these applications there are nearly always important secondary requirements to be met, e.g. that the glass should be easily workable, that its viscosity should not be influenced much by temperature, and so on. In

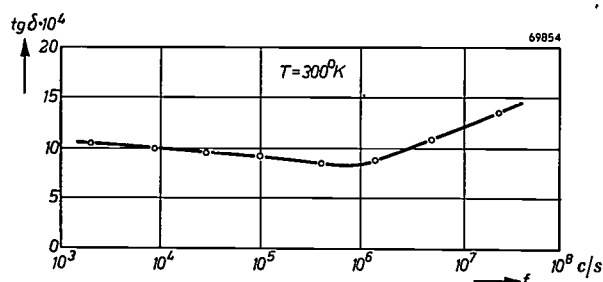


Fig. 12. Frequency spectrum at room temperature of the glass described under fig. 11. The great variety of network modifiers, including, i.a., also immobile Ca^{2+} and Pb^{2+} ions, causes the losses to be small and practically constant over a wide frequency range.

electrical engineering the glass is required to have a certain expansion coefficient to allow of its being fused onto metals such as steel, chromium iron, fernico and tungsten.

Most of the properties just mentioned can be obtained by using a high percentage of Na^+ ions in the glass, but then usually the losses are high. However, by making judicious use of the knowledge now acquired the losses can be appreciably reduced, while still retaining the other properties.

A good example is a glass of the composition 45% SiO_2 , 32% PbO , 5% CaF_2 , 4% Na_2O , 14% K_2O (weight percentages). Owing to its high contents of Na_2O , K_2O , CaF_2 and PbO , this glass softens at very low temperatures and has such a large expansion coefficient (120×10^{-7}) that it can be fused to iron, while at the same time it shows exceptionally small losses for the intermediate frequencies (see fig. 12; fig. 11 applies also for this glass). This is due to the addition of immobile ions

(Pb^{2+} , Ca^{2+} , K^+), while the presence of four kinds of network modifiers of different sizes promotes a good packing of the network. It is true that this glass has a relatively large R (viz. 2.64) — which gives it the low softening temperature — but this factor, which is unfavourable for the dielectric losses, is apparently sufficiently compensated by the other favourable factors mentioned above.

Investigation into the structure of glass with the aid of dielectric losses

Since small changes in the structure of glass often result in great changes in the dielectric losses, by applying the reverse process of reasoning it is often possible to draw interesting conclusions as to the structure of a glass from its dielectric losses.

A striking example of this is found in the dielectric properties of the borate glasses already discussed. The position of the separating line between the



Fig. 13. Variation of $\tan \delta$ at the frequency 1.5×10^8 c/s and at room temperature for a series of glasses of the composition 53.3% SiO_2 , 10.2% PbO , 4.5% CaF_2 and 32% $(Li_2O + MgO)$, as a function of the MgO percentage. As described under fig. 4, when some of the Li^+ ions are replaced by Mg^{2+} ions, $\tan \delta$ at first decreases, because the structure of the network is then more rigid. A further increase of the Mg^{2+} content reduces the rigidity of the structure and $\tan \delta$ increases. This does not continue along the same line, in contrast to fig. 4, since the higher the Mg^{2+} concentration, the more Mg^{2+} ions begin to act as network formers. Then they no longer contribute to the relaxation losses, so that ultimately $\tan \delta$ drops.

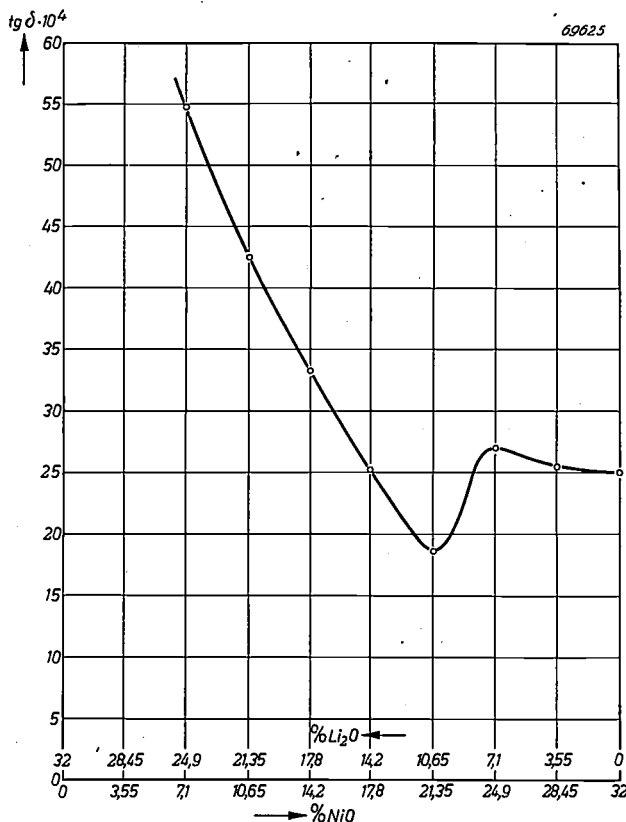


Fig. 14. This diagram relates to a glass analogous to that of fig. 13, but with the Li^+ ions replaced by Ni^{2+} instead of Mg^{2+} ions. The curve has the same trend as that in fig. 13.

accumulation region and the destruction region (figs 5, 6 and 9), predicted on theoretical grounds, can be experimentally determined, i.a. from measurements of the dielectric losses.

There is one other example that may be gone into more deeply. In article I it was stated that there are a number of cations which may occur both in a network-modifying and in a network-forming position. In a number of cases it is possible to see already from the colour of the glass how the ions behave. However, there are some ions (Mg^{2+} , Zn^{2+}) which do not colour the glass, and then the dielectric losses may serve as an indication how these ions are situated in the network. Particularly in the case of the Mg^{2+} ion most valuable information can thereby be obtained, for, as we have seen, this ion yields a rather considerable contribution towards the relaxation losses when occupying a network-modifying position, whereas it obviously contributes hardly anything at all to this kind of losses when it is in a network-forming position. Now it appears that, when the Mg^{2+} ion is present in a large concentration, it shows a tendency to occupy more and more the network-forming positions; such a network is poor in oxygen ions, so that the Mg^{2+} ion can compete successfully against the Si^{4+} ions. This effect is clearly demonstrated in fig. 13 (measure-

ments taken at 1.5×10^6 c/s), which is most instructive when compared with fig. 4. Whereas in the latter diagram the dielectric losses show the previously discussed deep minimum at the intermediate concentrations of Na^+ and Li^+ ions, the curve in fig. 13 (where two Li^+ ions have been replaced by one Mg^{2+} ion) shows that with small Mg^{2+} concentrations in the glass the curve for the losses follows an analogous trend, but with higher Mg^{2+} concentrations makes a turn and ends at a very low level, such due to the fact that part of the Mg^{2+} ions no longer contribute towards the losses. As a result of the Mg^{2+} ions changing over from network-modifying to network-forming positions, the value of R , going to the right, drops (in fig. 4 R is constant), and this in itself already leads to a further reduction of the losses.

This picture of the situation is illustrated by fig. 14, for the case where two Li^+ ions are replaced by one Ni^{2+} ion. This curve resembles very much that of fig. 13, and in this case it is indeed possible to conclude from the change in colour of the glasses examined that, going to the right, more and more Ni^{2+} ions occupy the network-forming positions.

Finally, it is to be mentioned that, measured at

$f = 10^{10}$ c/s, the replacement of two Li^+ ions by one Mg^{2+} ion yields a curve not showing any deflection but retaining the normal shape with a deep minimum. At these frequencies, where the vibration-losses play a part, it is of no consequence whether the Mg^{2+} ion occupies a network-forming or a network-modifying position. By way of illustration reference is made to fig. 8.

Summary. The dielectric losses in glass arise from three causes: conduction, after-effect and resonance. The various forms of losses are described, it being shown that the conduction losses are of importance only at very low frequencies and high temperatures. The after-effect losses can be divided into "relaxation losses", occurring at room temperature in the frequency range of 10^{-3} to 10^6 c/s and shifting towards higher frequencies as the temperature rises, and "deformation losses", which are approximately proportional to the frequency and are particularly of importance at low temperatures. The resonance or vibration losses occur mainly at high frequencies (10^{10} to 10^{13} c/s) and as the temperature rises probably shift towards lower frequencies. The influence of the chemical composition and the structure of the glass upon the various losses is investigated, special attention being paid to the borate glasses, in which an accumulation region and a destruction region are to be distinguished according to the composition. It is indicated how the losses can be reduced in certain frequency ranges, some examples being given. Finally, it is shown how the dielectric losses can be taken as indications of the changes taking place in the structure of a glass when its composition is modified.

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.

1981: P. C. vander Willigen: Grepn uit de ontwikkeling van het booglassen (Electro-techniek 29, 143-147, 1951, No. 8). (Some remarks on the development of arc welding; in Dutch.)

Two important questions in the field of electric arc-welding by hand are discussed in this article. The first part deals with the choice between direct current and alternating current. Their advantages and disadvantages are described, including the question of the open circuit voltage of the welding transformer. In the second part the different methods for welding of steel are examined from the point of view of shielding of the deposited metal against nitrogen and oxygen from the air. In this connection the use of hydrogen and its influence on the weld are also briefly discussed.

1982: N. Warmoltz: On the application of a Philips ionization gauge type of ion source in a mass spectrometer leak detector (Appl. sci. Res. B2, 61-65, 1951, No. 1).

A system of electrodes as in the Philips ionization gauge can function as ion source giving a diverging ion beam in which the ions have different velocities. When placing this source behind a slit, it is possible to focus all the ions of the same mass in a second slit by combining a magnetic field parallel to the slits with an electric field perpendicular to the slits and to the line joining the slits. Between the slits the ions describe a kind of epicycloidal trajectories. In this way it is possible to construct a simple leak detector.

1983: C. J. Bouwkamp and H. Bremmer: A note on Kline's Bessel-function expansion (Proc. Kon. Ned. Akad. Wetensch. Amsterdam A 54, 130-134, 1951, No. 2).

A certain combination of Bessel functions, viz. $J_\nu(\nu \sec \vartheta) Y_\nu(z + \nu \sec \vartheta) - Y_\nu(\nu \sec \vartheta) J_\nu(z + \nu \sec \vartheta)$, in which ν, ϑ, z are complex numbers, is expanded in a power series in the variable $1/\nu$. This series converges if $|\nu| > |z \cos \vartheta|$. The first three coefficients are given explicitly.

1984: J. H. van Santen and G. H. Jonker: Combinaisons ferromagnétiques du manganèse à structure pérovskite (J. Phys. Radium 12, 202-204, 1951, No. 3). (Ferromagnetic combinations of manganese with perovskite structure; in French.)

From investigations on the magnetic behaviour of ferromagnetic mixed crystals with perovskite structure of the type $(1-x)\text{LaMn}^{\text{III}}\text{O}_3 \cdot x\text{MeMn}^{\text{IV}}\text{O}_3$ (Me = Ca, Sr or Ba) it seems probable that the (indirect) exchange interaction between Mn^{IV} and Mn^{IV} is negative and that between Mn^{III} and Mn^{III} is positive. This is the first known example of a positive indirect exchange interaction in oxidic compounds.

1985: J. J. Went: Linear magnetostriction of homogeneous nickel alloys (Physica 17, 98-116, 1951, No. 2).

The magnetostriction of homogeneous Ni-alloys is investigated. The magnetostriction is measured as a function of the composition of the alloy, of the induction caused by an external magnetic field and of the temperature. A comprehensive table of data is given. Several general relationships for binary alloys have been found. The saturation magnetostriction λ_s at 0 °K can be predicted if the saturation induction I_s is known; λ_s at a higher temperature follows directly from the λ_s value at 0 °K and the relation I_s vs T (except for Ni-Mn-alloys). A single relationship has been found between λ and I for alloys except those containing Co or Mn. From all these data the magnetostriction for ternary alloys may be calculated.

1986: G. W. Rathenau and G. Baas: Grain growth in a texture, studied by means of electron-emission microscopy (Physica 17, 117-128, 1951, No. 2).

Grain growth and secondary recrystallization in rolled face-centered Ni-Fe-alloys has been studied by electron-optical means, an image of the activated hot metal surface being formed. Grain growth in an imperfect cubic texture proved to be a discon-

tinuous process. Almost all neighbouring grains invade quickly one grain or group of grains. Grain-boundary movement at low-energy boundaries occurs on a small scale. The high-energy parts of the boundary between a cubic crystal and its (near) twin proved to move in a direction parallel to the common (111) plane. The surface tension of the boundary between a secondary crystal and a cubic crystal which is to be absorbed is about twice the value corresponding to the surface between two cubic crystals.

R 160: H. Bremmer: The discharge of a series of equal condensers having arbitrary resistances connected in parallel (Philips Res. Rep. 6, 81-85, 1951, No.2).

This paper concerns the discharge of a series of condensers through a ballistic galvanometer if a variable resistor is connected in parallel to each condenser, in connection with the model described in R 159. The flow of current through the galvanometer is calculated with the aid of the operational calculus.

R 161: B. D. H. Tellegen and E. Klauss: Resonant circuits coupled by a passive four-pole that may violate the reciprocity relation (Philips Res. Rep. 6, 86-95, 1951, No. 2).

The system of two resonant circuits coupled by a passive four-pole that may violate the reciprocity relation is investigated. To obtain the maximum transfer for a given form of the resonance curve, the circuits must be equally damped and equally tuned. To obtain the maximum transfer with two circuits coupled by a passive four-pole satisfying the reciprocity relation, the circuits must generally be unequally damped and unequally tuned. When the coupled circuits are used as an interstage network in an amplifier, to minimize the influence of changes in valve capacitances on the shape of the resonance curve, the circuits should be equally damped and equally tuned. For symmetrical, flat-topped resonance curves the maximum transfer with non-reciprocal coupling four-poles, the ratios of the maximum transfer with reciprocal coupling four-poles to the transfer with reciprocal transfer with reciprocal coupling four-poles and to equally damped and equally tuned circuits are $(1 + \sqrt{2}) : 2 : 1$.

R 162: H. G. Beljers and W. J. van de Lindt: Dielectric measurements with two magic tees on shorted wave guides (Philips Res. Rep. 6, 96-104, 1951, No. 2).

A description is given of the method and calculations underlying dielectric measurements at microwaves with the aid of wave guides. It is shown that a „magic Tee” (a certain arrangement of four branches of rectangular wave guides at one junction) can be used both as a bridge and as means of establishing any complex waveguide impedance. By using these Tees dielectric measurements can be carried out with greater accuracy than in the conventional method with a standing-wave detector. Some details of the construction of the apparatus and the measurement are given. In conclusion a dielectric measurement of polystyrene is dealt with.

R 163: H. C. Hamaker and Th. Hehenkamp: Minimum-cost transformers and chokes, II (Philips Res. Rep. 6, 105-134, 1951, No. 2).

In the foregoing paper **R 150** the equations specifying a transformer design with minimum price, P , when the apparent power, VA , and the losses, W , are prescribed, were solved, and the chief characteristics of the resultant designs were discussed. This we shall call the P-W-class of designs. In the present paper two alternative classes of designs are considered, namely, (1) the P-M-class of designs giving minimum price for prescribed values of the power, VA , and the product, $M = BS$, of the peak magnetic flux density and effective electric-current density; and (2) the W-M-class of designs giving minimum losses, W , when VA and M are prescribed. The solutions obtained are casted in such a form that they are directly comparable *inter se* and with the P-W-solution of the foregoing

paper. When power and losses are prescribed the P-W-class of designs is always cheapest, but if instead of the losses we prescribe a value of B or of the specific dissipation of heat, v , per cm^2 of external surface, the other classes of designs may be preferable. It is shown that the choice of a design depends mainly on the transformer characteristics which are considered as the limiting factors; a choice between the three classes of designs is of secondary importance. These problems are discussed in detail. In the appendix, a set of tables is provided by means of which the P-M-designs or the W-M-designs can be rapidly obtained. A concluding section deals with the problem of a design such that the sum of the cost of the transformers and the cost of the electric power dissipated as heat during its life, is at a minimum. This problem can also be solved by the methods developed.

R 164: P. Schagen: On the mechanism of high-velocity target stabilization and the mode of operation of television-camera tubes of the image-iconoscope type (Philips Res. Rep. 6, 135-153, 1951, No. 2).

In this article the mechanism of the high-velocity target stabilization in the iconoscope and the image iconoscope is described. Some hypotheses are formulated for the mechanism of the image iconoscope, and equations are derived therefrom for the potential of target elements as a function of time and for the signal output of these tubes. This theoretical approach indicates that an increase in the secondary-emission coefficient of the target and in the target capacitance will result in better picture quality. Experiments confirm these predictions.

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 LINEAR ELECTRON ACCELERATOR

by D. W. FRY *).

621.384.622.2

Much of the work of contemporary nuclear physicists depends on the use of beams of high energy particles which are available from the particle accelerators constructed during the past two decades. Further development of techniques for producing such beams is directed towards higher particle energies as well as towards larger particle beam currents: higher energies are essential for discovering new types of reactions or elementary particles hitherto unknown, large currents are indispensable for certain investigations involving rather weak interaction effects, as for example neutron scattering.

The production of high energy electrons, in the millions of volts range, is of great importance also in another realm, medicine. The very penetrating X-rays emitted from targets under such electron bombardment give rise to dose distributions inside the human body which from a physical point of view should be much more favourable for effective therapeutic treatment than can be obtained with the 200 or 400 kV X-rays commonly used. Again, high X-ray intensities and, therefore, large electron beam currents are of prime importance for successful medical applications.

A very efficient means of meeting the demand for intense beams of electrons of several millions of volts energy, has been found in the linear electron accelerator. A series of articles on this development will be published in this Review within the next couple of years. The editors are greatly indebted to Mr. Fry, one of the most eminent authorities in the world in this field, for his permission to publish the following paper as the first in the projected series. Mr. Fry, who is at present Head of the General Physics Division of the British Atomic Energy Research Establishment, was at the head of the group of workers who built the very first travelling wave type linear electron accelerator.

*This paper is a slightly altered and shortened version of a lecture delivered at a symposium on Particle Accelerators at Delft in February 1951 **)* and gives a brief account of the fundamentals of the linear electron accelerator. The following articles of the series will describe i.a. the accelerators for 4 and 15 MeV respectively that were built by the Mullard Research Laboratory at Salfords (Surrey, England) under the direction of Dr. Bareford, in close collaboration with Mr. Fry's groups at Harwell and Malvern.

Introduction

Since about the time Cockcroft and Walton ¹⁾ disintegrated the lithium nucleus by bombarding it with alpha particles accelerated in a cascade generator to 800 keV, many ways have been proposed for accelerating particles to high energies. Some, such as the resonant transformer, the electrostatic generator, and the cascade generator, establish the full accelerating field between an insulated electrode and earth. They may be designed to have extremely useful properties such as high beam intensities and in the case of the electrostatic

and cascade generator a very narrow energy spectrum, but very formidable engineering problems are encountered if an attempt is made to design them for more than a few million volts energy.

To overcome the high potential insulation difficulty several alternatives were proposed, all of which in one form or another made use of the repeated action of quite low voltage fields upon the particles to be accelerated. Accelerators like the cyclotron, the betatron, the synchrotron and the synchrocyclotron which accelerate particles in cyclic paths under the action of magnetic and oscillating electric fields and which use frequencies up to a few tens of Mc/s form one broad group,

*) Atomic Energy Research Establishment, Harwell, England.

***) The complete lecture has been published in *De Ingenieur*, 64, O. 51, 1952 (No. 14).

whilst the methods in which the particles are accelerated in a straight line by either the continuous or intermittent action of (ultra) high frequency fields form another group. The latter principle was used successfully by Wideröe²⁾ in 1928 to accelerate potassium ions to 50 kV with a maximum applied voltage of 25 kV and a few years later by Sloan working first with E. O. Lawrence³⁾ and then with Coates⁴⁾ to accelerate mercury ions to 1.26 and 2.85 MeV respectively. Beams working with Snoddy⁵⁾ and with Trotter⁶⁾ also accelerated particles in a straight line using a different technique from Sloan et al. In their method co-linear electrodes were connected to the appropriate points of loaded transmission lines along which were transmitted voltage impulses. Protons and electrons were accelerated to energies of a few million volts by this method.

However, because of the success of the cyclotron for accelerating heavy particles, and the betatron for electrons, together with the lack of high powers at microwavelengths, which are essential for convenient configurations of linear accelerators, their development stopped until about 1945. By that time great advances had been made in the development of pulsed ultra high frequency power sources, due to the importance of such sources for radar and similar devices. Mean pulse powers of up to 2 MW's for several microseconds at repetition rates up to 500 pulses per second were available in the frequency range 150-3000 Mc/s. These high peak powers together with their associated new circuit techniques offered the possibility of the linear accelerator comparing favourably on technical and economic grounds with other methods of particle acceleration. The design of microwave linear accelerators using high powers and new techniques was started in several different laboratories.

Principle of travelling wave accelerator

The basic aim of all forms of linear accelerators is to establish suitably phased components of electric field along an axis down which the particles are to be accelerated, in such a way that the particle energy steadily increases. This will take place if the phase velocity of a progressive electromagnetic wave with a component of electric field in the direction of wave propagation can be controlled in a way which keeps the wave in phase with the particles being accelerated.

It may be seen that acceleration is possible in this way by reference to *fig. 1*. Let this represent the local variation, at any given instant in time, of the axial electric field of a wave progressing from

left to right, with peak field E . A charged particle at B (or D) is subject momentarily to the accelerating field E_m . If the particle at that instant moves to the right with a velocity much less (or greater) than that of the wave, it will continuously shift its position relative to the field distribution indicated, the force to which it is subject will rapidly change from accelerating to decelerating and vice versa, so that the mean energy imparted to the particle by the wave will be zero. If, however, the particle velocity at that instant is equal to the phase velocity of the wave, the particle will retain its relative position in the wave and it will be subject to the constant accelerating field $E_m = E \sin \theta_s$, as represented by the dotted horizontal line, θ_s being the phase of the particle relative to point O (zero field) in *fig. 1*. As the particle velocity will increase due to the accelerating force, its position in the wave can be maintained only if the phase velocity of the wave is made to increase likewise and at the same rate. The energy gained by a particle of charge e , after travelling a distance z in the constant accelerating field E_m , is $ezE \sin \theta_s$. Denoting the total energy (rest and kinetic) of the particle by ϵ and its rest mass by m_0 , its velocity v_p is given by the relativistic formula:

$$\epsilon = \frac{m_0 c^2}{\sqrt{1 - \left(\frac{v_p}{c}\right)^2}} = ezE \sin \theta_s + \text{const.} \quad (1)$$

(c being the velocity of light). Hence, this is the required dependence of the phase velocity $v = v_p$ of the wave on the distance z along the axis.

Under this condition, particles displaced from B (or D) find themselves in a field which produces either too much acceleration or too little, thus restoring them to the phase stable point B , around which the particles will oscillate. Likewise

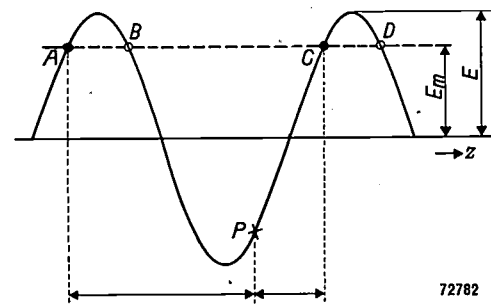


Fig. 1. Instantaneous values of electric field in z -direction, of wave travelling in this direction. The phase velocity of the wave increases in z -direction in such a way as to maintain a particle at B (or D etc.) in a constant accelerating field (E_m). Particles at B or D are in stable equilibrium, at A or C in unstable equilibrium.

particles at *A* (or *C*) are in unstable equilibrium. Mathematical analyses of the motion of particles trapped in a wave in this way have been carried out ^{7) 8)} and have shown there is damping of the amplitude of these phase oscillations with increasing energy both for slow and relativistic speeds and that generally the phase oscillations have a frequency much lower than that of the wave, almost ceasing to exist in the relativistic region. If particles are injected continuously into the wave with a velocity close to its initial phase velocity only those in the region from *A* to, say, *P* will be trapped in resonance with it and become bunched around *B*. Those starting with too large an initial displacement, i.e. from *P* to *C* in advance of the stable phase, will fall back beyond *A* and be lost.

The value of the phase θ_s of the position of stable equilibrium and, hence, the mean accelerating field acting on the particles oscillating around this position may be influenced by a suitable choice of the local variation of the phase velocity of the wave. With increasing mean accelerating field the trapping region decreases and it ceases to exist when this field equals the peak field. In the initial stages of an accelerator in order to trap a large number of particles it is advisable for the position of stable equilibrium to be well in advance of the wave crest. After trapping and bunching has been accomplished the equilibrium position, again by altering the phase velocity of the wave, may be moved nearer to the wave crest to obtain an increased rate of acceleration.

This description has ignored any radial motion of the particle. However, several workers ^{7) 9)} have shown using Maxwell's electromagnetic field equations for an axially symmetrical travelling wave, that in general there must be radial forces as well as an axial force acting upon the particles. In the region where phase stability exists the net radial force is defocusing, although as the particle velocity becomes close to the velocity of light the radial focusing force due to the circular component of the magnetic field becomes almost equal to the defocusing force due to the radial component of the electric field and the defocusing is only slight. Fortunately, radial stability may be restored by the application of a paraxial magnetic field of strength proportional to $1/\sqrt{V}$ where V is the kinetic energy of the particle.

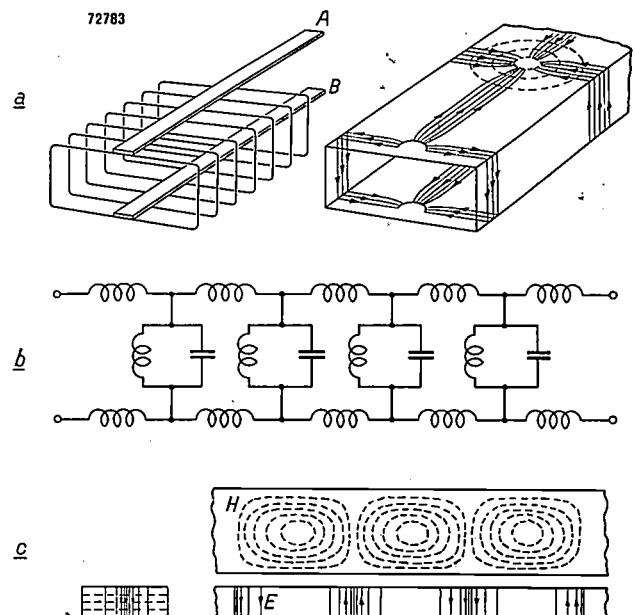
The corrugated waveguide

To accelerate particles in this way it was necessary to produce a progressive wave with an axial component of electric field and a phase velocity

equal to the particle velocity and therefore less than the velocity of light.

No well recognised method for producing a guided wave with these particular characteristics existed in 1945, although guided waves with a longitudinal component of electric field and a phase velocity greater than the velocity of light were well known by workers in the microwave field as one of many possible modes of wave propagation in a waveguide.

That waves may be propagated with other electric and magnetic field configurations than those which are set up on a two wire transmission line, i.e. the electric and magnetic field components mutually at right angles to each other and to the direction of propagation, may be illustrated by reference to a rectangular waveguide. This may be considered to be made up of two strips *AB* forming a twin wire transmission line continuously stub supported along both of its edges (*fig. 2a*). Such a



*Fig. 2. a) A rectangular waveguide propagating a H_{01} wave, with currents flowing in the walls as shown in the righthand picture, may be considered to be made up of a continuously stub supported twin wire transmission line *AB* (lefthand picture). b) Equivalent circuit of this transmission line. c) Pattern of electric field *E* (full lines) and magnetic field *H* (dotted lines) in the rectangular waveguide (*a*). The electric field has no component in the direction of wave propagation.

stub loaded transmission line may be represented by the equivalent circuit shown in *fig. 2b*, which may be regarded as a highpass filter whose cut-off frequency is the resonant frequency of the parallel tuned circuits, i.e. the frequency at which the stubs are $\lambda/4$ long. When the stubs occur very closely spaced, together they

form what is substantially a rectangular tube and the currents flowing in the transmission lines correspond to the currents in the top and bottom walls of a rectangular waveguide. These longitudinal currents fall to zero at the edges of the walls and there are no longitudinal currents in the side walls. The currents flowing in the stubs correspond to the transverse currents set up in the side walls when a H_{01} wave travels down the rectangular tube*). The electric and magnetic field patterns which are set up within the guide due to these currents flowing in the walls are shown in fig. 2c. The electric component of the field (full lines) is normal to the wide faces of the guide and at right angles to the direction of wave propagation, while there are both longitudinal and transverse components of the magnetic field (broken lines). This mode of propagation is of no interest for particle acceleration, since it has no axial component of electric field. However, there are others, the E modes, which are of interest.

light in cm/s and b the tube radius in cm. This is the cut-off frequency for the tube, which behaves like a conventional highpass filter. The phase velocity of these waves when propagated in a smooth-walled cylindrical tube may be shown to be greater than the velocity of light. To make them suitable for accelerating particles a means of controlling their phase velocity and reducing it to be less instead of greater than c has to be found. Several methods were proposed. The one which was adopted in the A.E.R.E. Laboratory and in the Microwave Laboratory of Stanford University uses metal irises, shown diagrammatically in fig. 3b. These irises form deep corrugations protruding from the wall of the cylindrical guide. For close spacings each corrugation may be regarded as a radial parallel plate transmission line. When their depth is less than a quarter of a wave they provide an inductive loading in the guide wall and by analogy with an inductively loaded transmission line it is to be

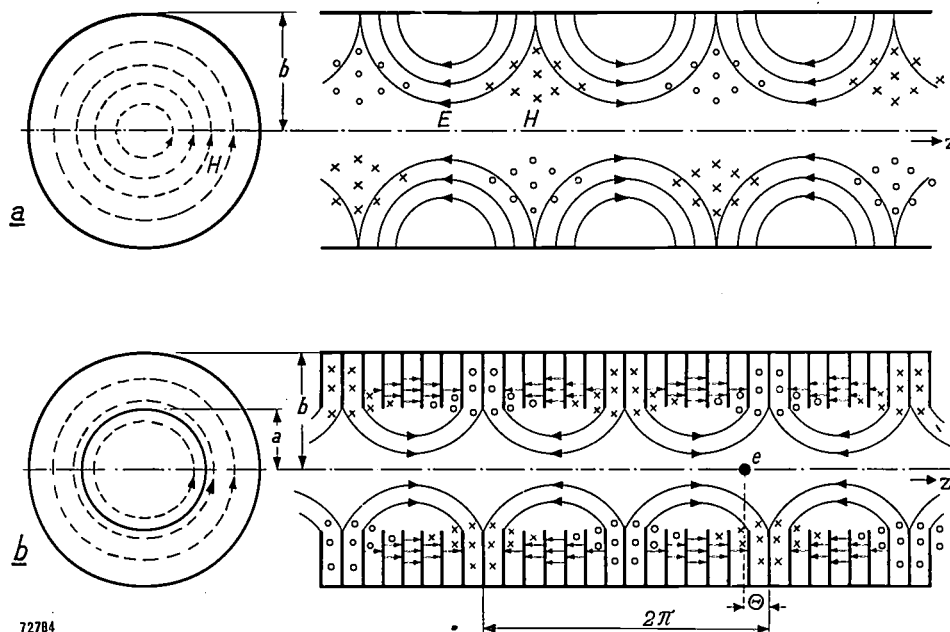


Fig. 3. a) Field pattern of E_{01} wave in circular waveguide of diameter b . Full lines electric field. The lines of magnetic force are circles around the axis.
b) Field pattern of E_{01} wave in corrugated circular waveguide. e = typical position of particle experiencing acceleration.

Stratton¹⁰) for instance shows that in a smooth-walled guide of circular cross-section electromagnetic waves having the field patterns shown in fig. 3a may be propagated if their frequency (in c/s) is greater than $0.38 c/b$, where c is the velocity of

expected that the velocity of propagation of the waves passing along the tube will be decreased.

A careful study of the equations governing the field existing within such a structure was first made by Cutler¹¹), who obtained solutions of them which were satisfactory for many purposes by assuming that the corrugations were thin and closely spaced compared with a wavelength. These solutions confirmed that wave propagation in such

*) Readers may refer to the articles by W. Opechowski, Philips techn. Rev. 10, 13-25 and 46-54, 1948, which contain a somewhat more elaborate discussion of waves in waveguides of different shapes. Ed.

a structure could take place and that by *suitable control of the corrugation depth the phase velocity v_p for a given frequency could be made equal to or less than the velocity of light*; in fact, a specific value could, within limits, be obtained at any given point of the guide. They also showed that the fields exhibit some rather special characteristics when $v_p = c$. In particular the accelerating field is uniform over the open part of the tube and (in Giorgi units) is at any cross-section given by

$$E = (480 W)^{\frac{1}{2}} (\lambda/\pi a^2), \quad \dots \quad (2)$$

where W is the power flowing through the cross section, a is the radius of the iris aperture and λ the wavelength. For the same input power to guides of different radii the accelerating field is inversely proportional to the square of the inner radii, and as the inner radius decreases the fields inside the corrugations tend to those in a cavity resonator supporting an E_{010} mode.

As the power travels along the guide it is attenuated by copper loss in the plane dividing walls of the corrugations. The number of irises should therefore be made as few as possible without seriously reducing the efficiency of the guide for accelerating particles. Walkinshaw¹²⁾ has shown that for a travelling wave accelerator the best performance is obtained when there are 3-4 corrugations per wavelength. With a larger number the copper loss is unnecessarily high, with fewer, i.e. as the corrugation spacing increases, the accelerating field for a given power flux decreases. This decrease occurs because the field between corrugations becomes less uniform so that higher harmonic components of the wave are set up. These carry a proportion of the power which is no longer available for acceleration.

With two corrugations per guide wavelength the travelling wave form of accelerator is no longer possible. The irises occur at half wavelength spacings and each corrugation is then in resonance with the frequency of the radio frequency power being fed into it. A standing wave instead of a travelling wave pattern builds up in the guide and it becomes a multiple resonator with a phase change of 180° between successive corrugations. This is the π -mode of resonance often used in magnetrons. When the periodic structure is resonant it is no longer possible to transmit power through it.

A resonant structure such as this with a standing wave field pattern set up in it provides an alternative form of linear accelerator, for the standing wave pattern may be regarded as being made up of two travelling waves moving in opposite directions. Particles injected into such a field with a velocity equal to that of one of the component waves would experience similar forces acting upon them to particles in a travelling wave accelerator for the effect of the reverse wave upon them must on a time average be zero. It has been claimed⁷⁾ that the standing wave method has an important

technical advantage over the travelling wave method despite the wastage of power in the reverse wave. This is its ability to act as the frequency stabilising cavity for the radio frequency oscillator driving it. However, this has never yet been demonstrated experimentally to be an important advantage.

Use of corrugated waveguide for travelling wave accelerator

Several papers have now been published^{8) 9) 13)} on the properties of corrugated guides suitable for travelling wave accelerators. They give theoretical data of the axial fields obtainable in metal corrugated guides when the phase velocity is made equal to c for different iris apertures and taking into account attenuation of power in the guide. Actually particles are injected into the waveguide at a velocity much less than c : in fact for electrons an initial velocity of about $0.4 c$ is chosen, but after "surf-riding" on the wave over only a short distance, velocities very near c are attained, further acceleration virtually causes only an increase in mass of the electron (at 2 MeV energy, the electron velocity is already about $0.98 c$).

Because of the attenuation of the power, the accelerating field E according to formula (2) decreases along the guide. The total energy of a particle at the end of the guide is given by the integral of E over the given length (cf. form. 1). Harvie⁹⁾ has published graphs showing the integrated field (particle energy) in a series of guides of differing inner radii fed with 2 MW's peak power on a wavelength of 10 cm for guides with 10 corrugations per wavelength. This is shown in *fig. 4a*. The independent parameter is the inner radius expressed as a/λ where λ is the free space wavelength of the radio frequency power being fed into the waveguide. The envelope of the family of curves, the top chain line, shows the maximum attainable energy (i.e., assuming the particles to travel at $\theta_s = 90^\circ$) for any given length of accelerating tube, each length corresponding to a definite value of a/λ . The limitation imposed on each curve where it flattens off is entirely due to attenuation. However, other more serious limitations to the maximum energy attainable in an accelerator may often be present, such as those imposed by frequency and temperature variations and also by dimensional errors arising in manufacture.

The importance of the frequency stability of the radio frequency source is to be seen from the remaining dotted curves in this figure. They show the limit of energy attainable with a guide of given a/λ when it is assumed that the maximum permissible phase error of the particles in the wave is

$2\pi/10$ radians. In an ideal accelerator the wave and particles would remain in step along the whole length of the accelerator irrespective of the radio frequency. In a corrugated waveguide this does not happen because the energy stored in the corrugations is a rapidly varying function of frequency which gives it the characteristics of a highly dispersive medium. If the waveguide is designed to have

the figures are sufficient to show that the waveguide dimensions must be determined with a high degree of precision. In fact with an intolerably high degree for long accelerators. Fortunately it is possible to construct a long guide of a number of sections joined together by short sections which may be used for phase correction. In practice it has been found that a guide 2 metres long with $a/\lambda = 0.2$

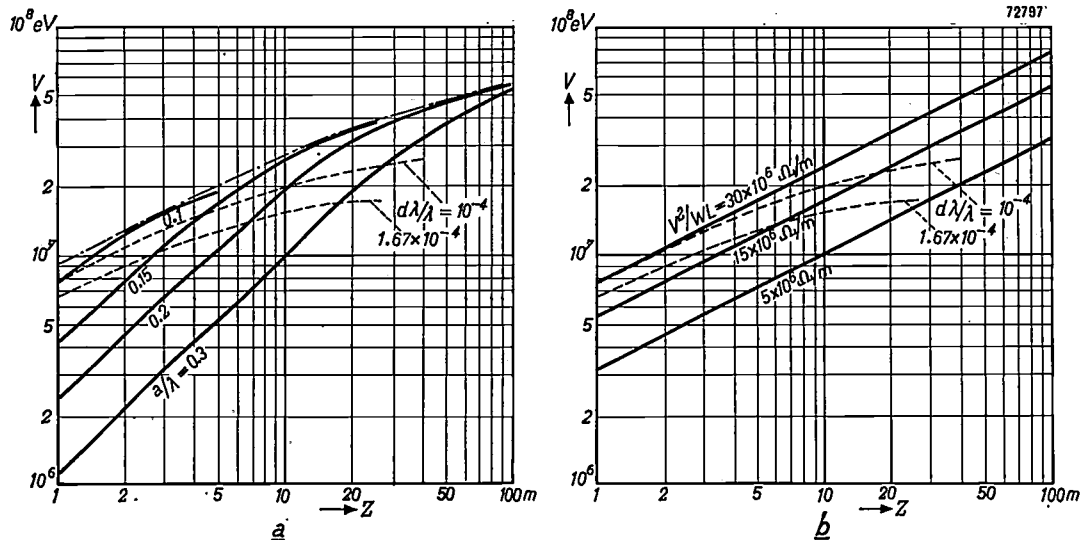


Fig. 4. a) Maximum particle energy V obtainable in a corrugated waveguide of given length Z , for different inner radii a of the irises. The curves are calculated for a guide containing 10 irises per wavelength, a phase velocity $v_p = c$ (the stable position of the particles — electrons — being located at the wave crest, $\Theta_s = 90^\circ$), wavelength $\lambda = 10$ cm, input power 2×10^8 W. The flattening off of the curves is due to the attenuation of the wave travelling in the guide. The intersection of each full line with one of the broken lines indicates the guide length Z at which with a given frequency stability $d\lambda/\lambda$, the phase error attains the assumed limit of 36° .

b) Constant "efficiency" contours ($V^2/WL = \text{const.}$, see below) graphed together with the frequency tolerance curves ($d\lambda/\lambda = \text{const.}$) from (a). With a given frequency stability high efficiencies can be obtained by using a short length $Z = L$ and a small a/λ value.

the correct phase velocity at a given frequency, then at any other frequency the particles being accelerated will no longer be in the correct phase relationship with the wave. It may be seen that frequency variations of a few parts in 10^4 , i.e. about the best stability that can be obtained in practical cases, impose severe limitations on the permissible length of guide.

A similar effect is to be expected due to temperature variations of the waveguide. For the coefficient of expansion of copper is about 1.65×10^{-5} , thus a change of 1°C in guide temperature (changing the radius b) corresponds to a $d\lambda/\lambda = 1.65 \times 10^{-5}$; or a temperature variation of $\pm 10^\circ \text{C}$ is equivalent to a wavelength tolerance curve of $d\lambda/\lambda = 1.65 \times 10^{-4}$. The effects of temperature variations are easier to overcome than frequency changes for they occur more slowly and may be compensated by small changes in the operating frequency. Nonetheless,

and ten corrugations per wavelength may be constructed to have a phase error within approx. 18° of that predicted by theory. For a guide with this value of inner radius this seems to be approximately the right section length.

It will be seen in fig. 4a that for a given power input the acceleration per unit length increases rapidly and the maximum usable length decreases with decreasing inner radius. When it is required to accelerate particles to a given energy it is relevant to consider whether a long guide of large inner radius or a shorter guide of smaller inner radius should be used.

In deciding this it is convenient to consider the relative waveguide "efficiency" as defined by the ratio

$$\frac{(\text{energy})^2}{\text{peak input power} \times \text{guide length}} = \frac{V^2}{WL} \frac{\text{megohms}}{\text{metre}},$$

when V is in MeV, W in megawatts and L in metres.

Since this is a function both of power and length it is an indication of economy. Also it is almost the same as the effective series impedance of a unit length of accelerator and so may be used to make a rapid assessment of the characteristics of any high energy accelerator, which may be considered

The power attenuation measured in guides made in this way was approximately twice the theoretical value for pure copper. Later guides made for a 10 MeV accelerator using a silver soldering method have had considerably greater mechanical strength than the A.E.R.E. ones and the copper loss was

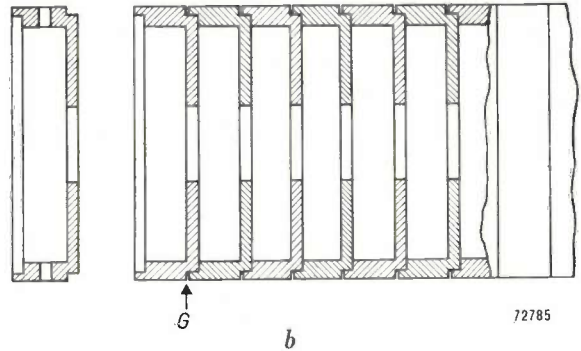
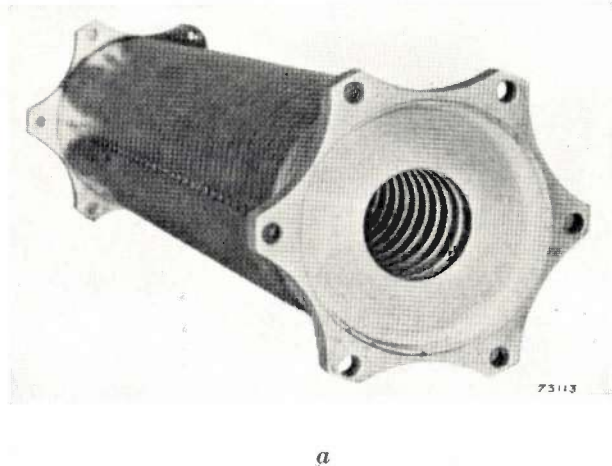


Fig. 5. a) 40 cm section of corrugated circular waveguide, made at A.E.R.E. b) Construction of the guide shown in (a). To the left a single machined element, to the right several assembled elements (partly sectionalised). The push fitting elements are clamped together and the 0.010" gap G is filled with soft solder.

to be made up of a number of component accelerators. Constant "efficiency" contours may also be graphed together with the frequency tolerance curves as has been done in fig. 4b. These contours are straight lines of equal slopes, with the "efficiency" proportional to the square of the intercept on the ordinate. They are cut by the frequency tolerance curves in a similar way to that in which the a/λ curves are cut in fig. 4a. It may be seen from figs. 4a and b that for high "efficiencies" it is necessary to use small a/λ 's and short lengths.

Waveguide construction and performance

Mullett and Loach¹⁴) have described in detail the methods of feeding, and measured characteristics of corrugated guides made at A.E.R.E. Fig. 5a shows a photograph of a section 40 cms long with an $a/\lambda = 0.2$ and 10 corrugations per wavelength, and its method of construction is shown in fig. 5b. Each corrugated element was machined from pure H.C. copper and made self locating with its adjacent elements by spigots. Gaps of 0.010" were left on the outside of the guide between elements, which were later filled with soft solder to provide some mechanical strength, but the obtaining of good conductivity across the joint at the bottom of the corrugations relied upon pressure contact.

only approximately 25% greater than that predicted by theory.

To investigate the accuracy of the theory relating phase velocity with the various guide parameters such as inner radius, depth of corrugation, pitch of corrugation, etc. a number of guides were made up in which these parameters had been suitably varied.

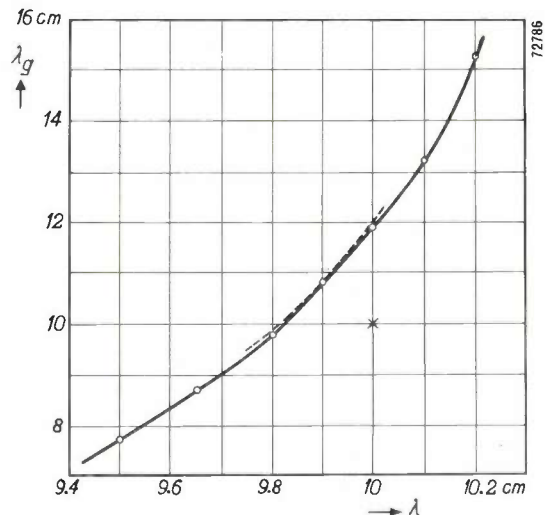
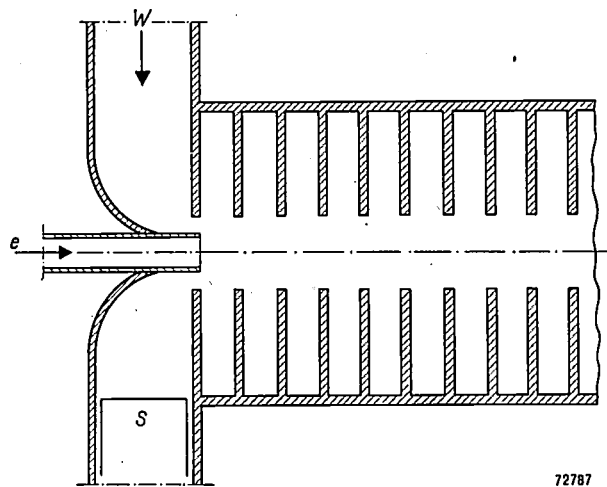


Fig. 6. Wavelength λ_g in corrugated circular guides as a function of frequency (wavelength λ in free space). The dots indicate values measured for guides with 2 cm pitch, $a/\lambda = 0.2$ and outer iris radius b calculated on the elementary theory to give $\lambda_g = 10$ cm for $\lambda = 10$ cm (x). The dotted line is calculated from a more refined theory by Walkinshaw⁸).

The phase velocity in each guide was measured as a function of frequency by shorting it at one end, feeding power into it, and measuring the standing wave pattern set up in accordance with the relationship $v_p = c\lambda_{\text{guide}}/\lambda_{\text{air}}$. A typical measurement of λ_{guide} versus λ_{air} is shown in fig. 6 together with the agreement obtained with the approximate theory, and also a more refined one by Walkinshaw⁸) which took account of the field distribution at the mouth of the corrugations.

Two methods were developed for feeding both the radio frequency power and the particles into cylindrical corrugated guides. The first used a tapered coaxial line of suitable dimensions for the magnetic field component in it and in the first cavity of the corrugated guide to be impedance matched. The second method, described as a door-knob feed, was developed by Mullett. It is shown diagrammatically in fig. 7. This design is empirical, but it has considerable advantages over the first method both as regards compactness and ease with which it is matched into the waveguide.

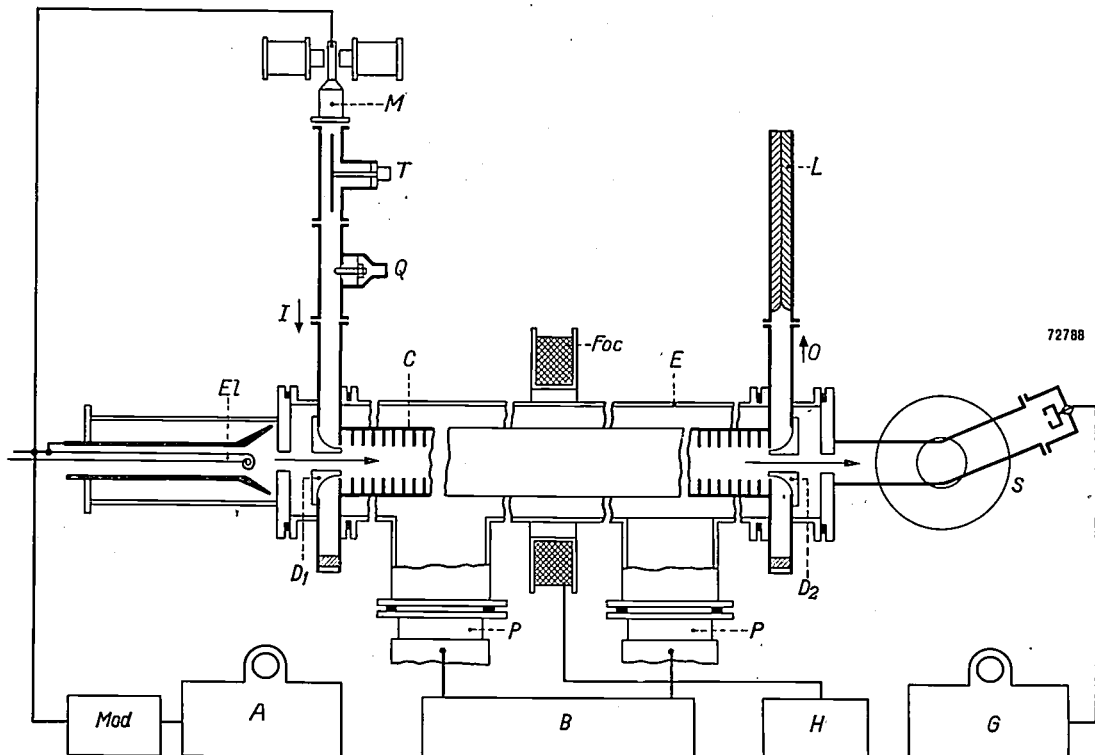
In the above, the principle of the travelling wave



72787

Fig. 7. "Door-knob" feed for the corrugated waveguide. The radio frequency power W is fed through a rectangular waveguide $3'' \times 1\frac{1}{2}''$, the opposite section of rectangular guide closed by an adjustable piston S enabling the exact matching of impedances. Particles (electrons) are injected from e .

linear accelerator and its mode of operation were outlined with consistent reference to "particles" in general. It must now be stated that considerable practical difficulties are encountered in applying



72788

Fig. 8. Block schematic diagram of the 3.5 MeV linear electron accelerator built at A.E.R.E. C tapered corrugated waveguide in vacuum envelope E , El electron gun, I radio frequency input, D_1 door-knob feed, D_2 door-knob exit, O radio frequency output, L steel load taking up remainder of wave power; M magnetron with modulator Mod ; T and Q line lengthener and probe for adjusting the magnetron operating frequency, A synchronising, pulse monitoring, safety interlock, remote control; Foc one of 15 focusing coils, H focusing current controls, B mechanical rotary backing pump, vacuum pressure monitoring equipment and automatic control; S electron spectrometer with deflection coils and Faraday chamber, G spectrum analyser control and display unit.

the principle to heavy particle acceleration, so that the remaining part of the article in which a specific machine is to be described only deals with the acceleration of electrons.

The A.E.R.E. 3.5 MeV electron accelerator

The principal characteristics of this accelerator have been described by Fry et al.¹⁵⁾ *Fig. 8* is a block schematic diagram of it. It is designed to work with a peak input power of 2 MW's in 2 μ sec pulses obtained from a magnetron. The power is fed through a $3'' \times 1\frac{1}{2}''$ rectangular waveguide to the input door-knob feed and then into 2 metres of corrugated guide. Unused power, reaching the end of the corrugated guide again passes via a second door-knob feed into a rectangular waveguide and is absorbed in a dummy load. The accelerating tube is mounted inside a vacuum envelope. This envelope is in turn surrounded by focusing coils to provide the steady axial magnetic field necessary for neutralizing the radial defocusing forces acting

on the electron beam as mentioned above. Electrons are injected from a diode gun of the Pierce type *) which is pulsed to 45 kV by the modulator which also drives the magnetron. The electrons are injected with a velocity of $0.4c$ into a wave of the same initial phase velocity. By a slight tapering of the first part of the corrugated wave guide, the phase velocity of the wave is matched to the increasing velocity of the electrons. The equilibrium position for accelerated electrons is at first 45° behind the field zero but in the first 40 cms length, by suitably adapting the tapering of the guide, this position is shifted up 10° , to 55° , to take advantage of a higher accelerating field after the electrons have become well bunched. The residual gas pressure measured by ionisation gauges mounted on the vacuum envelope above the pumping ports is 2×10^{-5} mm Hg under normal running conditions. The electrons arriving at the end of the corrugated guide

*) Cf. Philips techn. Rev. 13, 216, 1951 (No. 8).

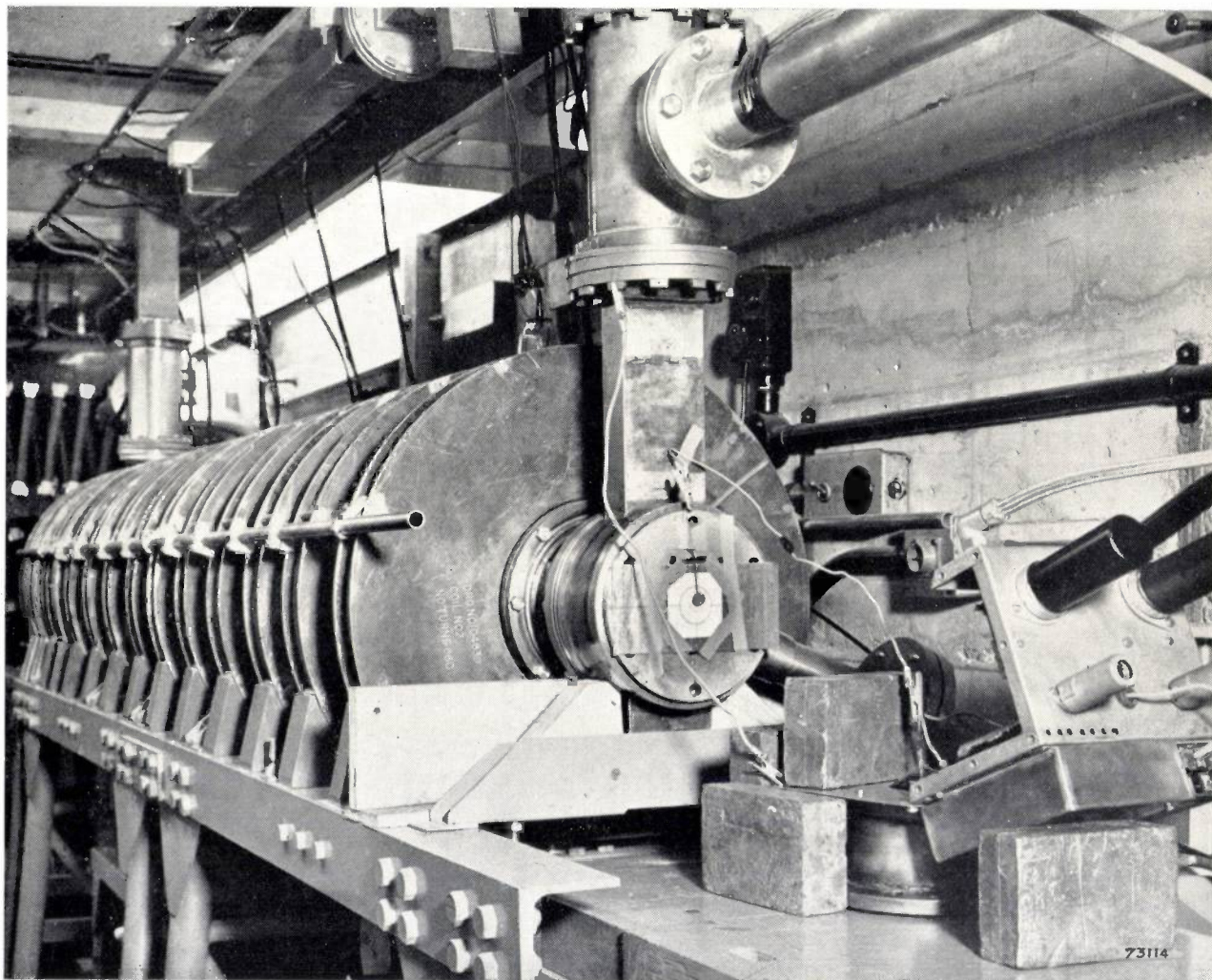


Fig. 9. Photograph of the output end of the 3.5 MeV electron accelerator.

are allowed to strike a lead target, their energy being converted into X-rays. A photograph of the output end of the accelerator is shown in *fig. 9*.

This particular guide was designed for no (or very small) beam loading. When a beam of electrons is fed into the guide for acceleration the effective attenuation in the corrugated waveguide increases since appreciable energy is now being transferred to the electrons. Hence, except at the input end, the radio frequency power flow and consequently the field diminish with increasing beam loading. This effect is partially compensated in the bunching region by the electrons tending to move closer to the peak of the wave, i.e. into a region of higher field. Over the main portion of the accelerator, however, the electrons remain at a constant phase with respect to the wave with the net result that the electron energy attained decreases with increasing gun filament (or emission) current.

Preliminary operating data as a function of the gun filament current were published but later and rather more reliable curves obtained with improved measuring techniques are shown in *fig. 10*. Taken at an operating frequency of 2997.6 Mc/s with a peak input power of 2.1 MW's and a pulse repetition rate of 50 pulses per second the curves show electron energy and mean beam intensity, the mean power in the beam, and the X-ray intensity in roentgens/min measured at a distance of 1 metre from the target. With the operating conditions specified the mean input radio frequency power to the system was 210 watts and the highest mean power measured

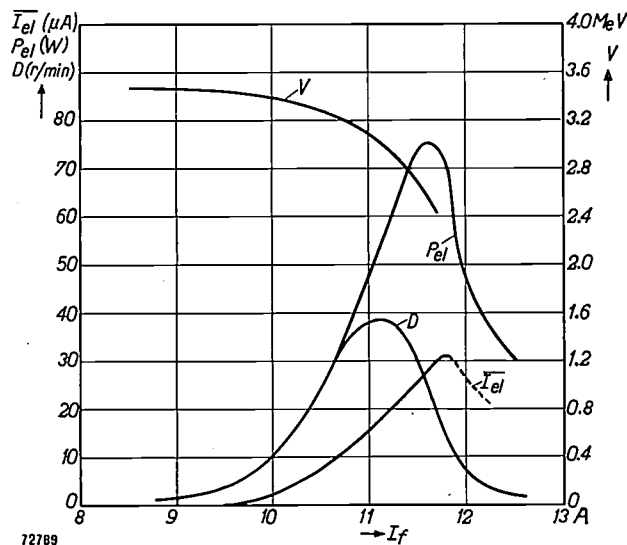


Fig. 10. Operating characteristics of the 3.5 MeV electron accelerator. I_f gun filament current, V energy of the electrons, I_{el} mean beam current, P_{el} beam power (dissipated in target), D X-ray dose rate in röntgens/min at 1 metre distance from lead target. Frequency 2997.6 Mc/s, peak input power 2.1 MW, pulse length 2 μ sec, repetition rate 50 pulses per second (mean input radio frequency power 210 W).

in the beam 76 watts, i.e. an efficiency of 36%, at a beam energy of 2.6 MeV. The highest mean current measured was 20 μA 's, which with the duty cycle of 10 000:1 corresponds to a mean current in the pulse of 300 mA. Typical operating conditions for the machine are: energy 3.25 MeV, 10 μA 's mean current and an output of 31 röntgens/min at 1 metre at 50 pulses per second, although the repetition rate is often increased up to 250 pulses per second. The energy spectrum when the operating conditions are correctly adjusted is a clean one about the mean energy with a full width at half intensity of approximately 5%.

Linear accelerators with feedback

The discussion of *fig. 4* may be summarized in the statement that the most "efficient" corrugated waveguides are those with small a/λ and that the maximum "efficiency" for a guide of a given inner radius occurs when it is long enough for most of the radio frequency power not transferred to the electron beam to be dissipated in the walls with little passing out into the dummy load. These two conditions are also those which demand a very high degree of frequency stability from the radio frequency source if particle acceleration is to be maintained along the whole guide length. The designer of a linear accelerator of the type described may therefore find himself presented with the choice of making a short efficient accelerator which is critical to adjust and operate, or a longer one which is less efficient and correspondingly less critical to set up. This problem is of particular importance when considering the application of the linear accelerator for clinical use. In such an application, the beam intensity must be adequate to allow the treatment times to be kept short even after heavy filtration and collimation of the X-ray beam, while on the other hand the instrument must be engineered in a form which makes the setting up of a patient easy, thus a long and clumsy accelerator would be most undesirable.

A method of improving the "efficiency" of corrugated guides of length short compared with the attenuation length and thus avoiding the need to make the choice described, has been worked out by Harvie and Mullett¹⁶). It consists of applying radio frequency power feedback in the manner shown diagrammatically in *fig. 11*. The directions of power flow are shown by the arrows, power from the output end of the accelerator being fed back into a suitable bridge circuit where it is combined with the incoming power from the magnetron. It will be noted that the power at the beginning

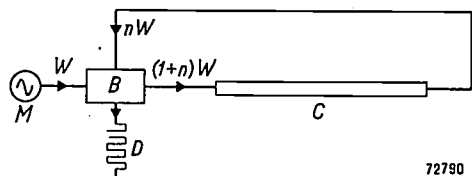


Fig. 11. Schematic diagram of radio frequency feedback circuit. C corrugated guide, B bridge circuit, M magnetron with output power W , D absorbing load (this is necessary only for the build-up period of the power in the guide during every pulse).

of the corrugated waveguide (usually called the circulating power) is then greater than that entering from the magnetron. The novel feature of this system is the radio frequency bridge which is required to have two special characteristics. It must be designed in such a way that all the power is made to flow through the bridge in the proper direction, and the input impedance as seen by the magnetron must be independent of the amount of power fed back from the end of the accelerator. This is necessary in order that the input impedance shall remain constant while the power is building up in the system. The design of such a bridge for the special case when (in the steady state) the power entering the bridge from the feedback arm is equal to the power entering from the radio frequency source has been studied extensively on microwavelengths, the bridges being known as "Magic T's" and "Rat Races"¹⁷⁾. Calculations of the "efficiency" V^2/WL of a fed back accelerator have been made and it is found to increase as the length of the accelerator is decreased from the length for maximum "efficiency" without feedback. The relative "efficiency" in terms of the accelerator length expressed as a fraction of the optimum length is shown in fig. 12 for the case of no feedback and feedback applied with and without loss in the return path. It may be seen clearly that accelerators short compared with the attenuation limited length may be constructed without loss of "efficiency". This enables increased dimensional tolerances to be permitted on the wave-

guides and a lower frequency stability from the radio frequency source to be accepted.

Advantage has been taken of this advance in technique in the two different forms of linear accelerator now being developed in Great Britain for clinical use. They are to give 4 and 10 MeV respectively. A number of the lower energy machines are to be built for distribution and use at large medical centres but at present only one is being constructed at the higher energy. Basically both designs are similar to the one described above, but

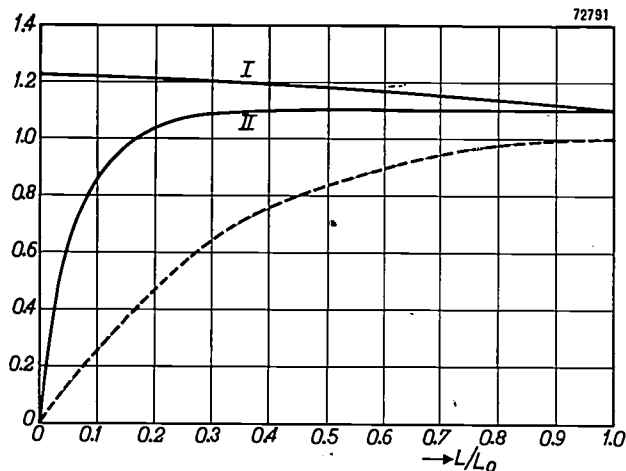


Fig. 12. "Efficiency" V^2/WL of accelerator with feedback (full lines) and without feedback (dotted line) in terms of the accelerator length expressed as a fraction of the optimum length L_0 derived from fig. 4. Curve I: no loss in return path; curve II: loss in return path equal to 5% of that in corrugated guide of length L_0 .

they differ markedly in many of their physical characteristics and particularly in the ways in which they are mounted.

Important contributions to the work reviewed in this paper have been made by R. B. R-S-Harvie, L. B. Mullett and W. Walkinshaw of the A.E.R.E. Linear Accelerator Group. Acknowledgement is also made to the Director of the A.E.R.E. for permission to publish this article.

Literature:

- 1) J. D. Cockcroft and E. T. S. Walton, Proc. Roy. Soc. A 129, 477, 1930; Ibid. 136, 619, 1932.
- 2) R. Wideröe, Archiv Elektrotech. 21, 387, 1928.
- 3) D. H. Sloan and E. O. Lawrence, Phys. Rev. 38, 2021, 1931.
- 4) D. H. Sloan and W. M. Coates, Phys. Rev. 46, 539, 1934.
- 5) J. W. Beams and L. B. Snoddy, Phys. Rev. 45, 287, 1934.
- 6) J. W. Beams and H. Trotter, Jr., Phys. Rev. 45, 849, 1934.
- 7) J. C. Slater, Rev. Mod. Physics 20, 473, 1948.
- 8) W. Walkinshaw, Proc. Phys. Soc. 61, 246, 1948.
- 9) R. B. R-S-Harvie, Proc. Phys. Soc. 61, 255, 1948.
- 10) J. A. Stratton, Electromagnetic Theory, McGraw-Hill, New York 1941.
- 11) C. C. Cutler, Bell Telephone Labs. Report MM-44-160-218, 1944.
- 12) W. Walkinshaw, T. R. E. Report AE 1005, 1947.
- 13) E. L. Ginzton, W. W. Hansen and W. R. Kennedy, Rev. sci. Inst. 19, 89, 1948.
- 14) L. B. Mullett and B. G. Loach, Proc. Phys. Soc. 61, 271, 1948.
- 15) D. W. Fry, R. B. R-S-Harvie, L. B. Mullett and W. Walkinshaw, Nature 162, 859, 1948.
- 16) R. B. R-S-Harvie and L. B. Mullett, Proc. Phys. Soc. 62B, 270, 1949.
- 17) L. D. Smullin and C. G. Montgomery, Microwave Duplexers, McGraw-Hill, New York 1948.

Summary. When particles are injected into an electromagnetic wave of phase velocity equal to the particle velocity and with a component of electric field in the direction of wave propagation, the particles will gain energy from the wave.

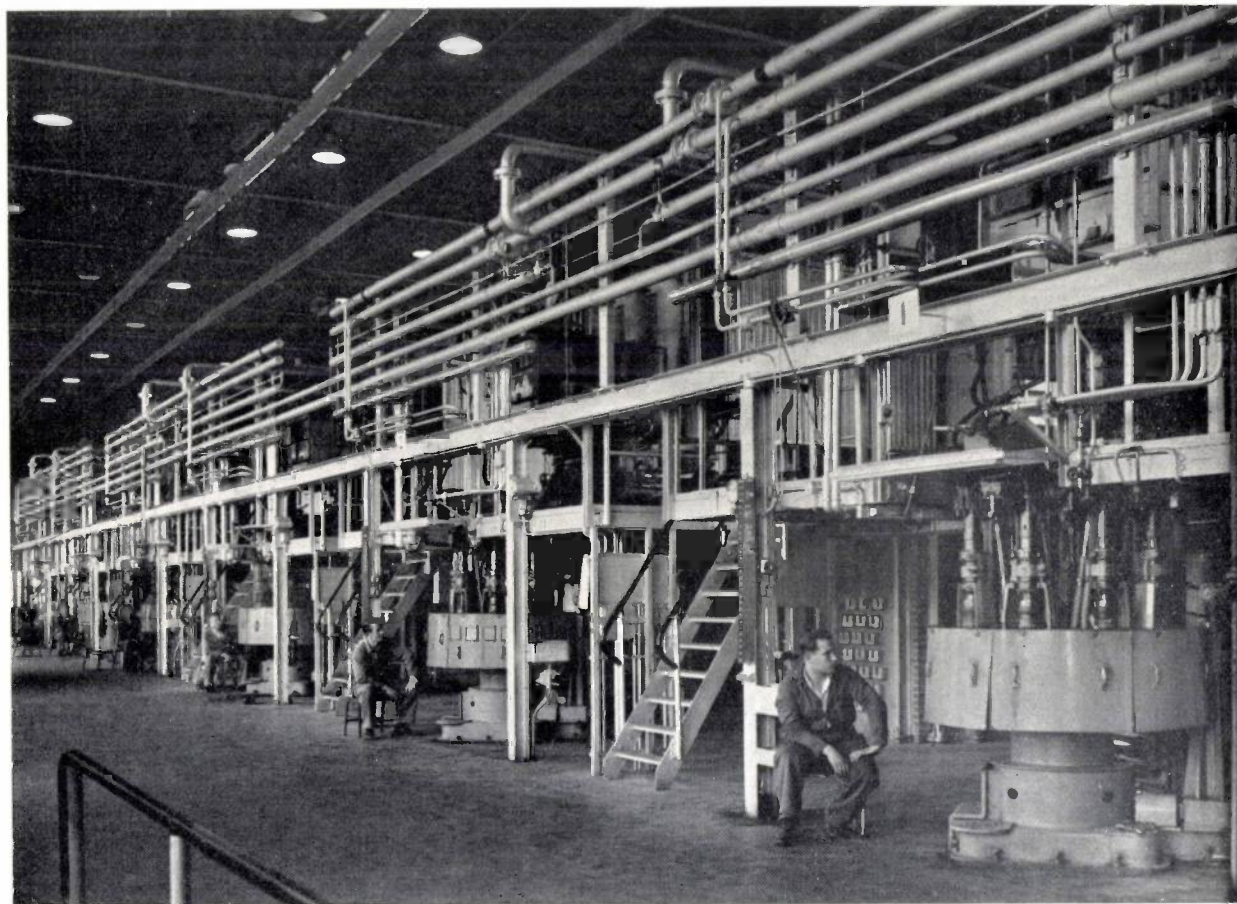
Electromagnetic waves of the characteristics required may be obtained in a cylindrical waveguide loaded by irises (corrugated waveguide). By a slight tapering of the guide diameter the phase velocity may be controlled in such a way

as to keep the waves in step with the particles being accelerated. Under these conditions the particles take up positions of stable equilibrium in the wave and may be accelerated to very high energies, depending on the power of the wave entering the guide, the attenuation of the wave progressing in the guide, and the ratio of inner radius of the corrugations to wavelength. This method of accelerating particles, while known before 1939, became of great interest with the advent of radio frequency sources yielding very large pulse powers on microwavelengths, allowing very high particle energies to be attained with convenient accelerating guide configurations.

In this article the conditions for acceleration are described and some of the fundamental properties of this form of travelling wave particle accelerator outlined. Theoretical and practical limits imposed on the attainable energy are dis-

cussed and the "efficiency" is considered. The possibilities of the linear accelerator are illustrated by reference to the characteristics of a 2 metre long electron accelerator now in operation at the Atomic Energy Research Establishment, Harwell, England. In this instrument electrons are accelerated from 45 keV to approximately 3.5 MeV. The operating wavelength is 10 cm. The radio frequency power is obtained from a magnetron in pulses of 2 MW mean value and 2 microseconds pulse length, the maximum pulse recurrence frequency being 250 pulses per second. Mean pulse currents of the beam of accelerated electrons as high as 300 mA have been measured. In machines of more recent design the principle of feedback of the unused radio frequency power has been applied, enabling the acceleration of electrons to high energies in short guides without the inconvenience of too strict frequency stability requirements.

THE PHILIPS BULB BLOWING MACHINE



Glass bulbs form part of the "food" devoured daily in great quantities by incandescent lamp and radio valve factories. They were blown originally, as were so many other glass articles, by means of the human breath, and even up to a few years ago this very old process played its part in the mass production of standardised lamp and valve bulbs ¹⁾.

¹⁾ See Philips Techn. Rev. 6, 74, 1941.

Nowadays such bulbs are entirely mechanically produced. The photograph shows a number of bulb blowing machines, standing in a bay of the Philips Glass Factory in Eindhoven. Each machine has its own glass furnace, located on the gallery, and fed from the rear with the raw materials for the glass. Below the front portion of each furnace is an opening down through which the glass flows. There is, however, no continuous stream of glass, since immediate-

ly below the opening is a kind of scissors which cuts the stream of glass into portions of a pre-determined weight. Periodically, therefore, a portion of glass falls down and is caught by a continuously rotating turntable, some of which are to be seen in the photograph. The glass is blown on the turntable into a bulb, which is just completed after one revolution of the turntable. The bulb is then ejected and carried by means of a conveyor belt (not visible in the photograph) through an annealing oven to remove strain; via a second conveyor belt, on which it cools, the

bulb reaches the inspector who sorts the bulbs according to quality.

The difficulties encountered in developing such an automatic machine lie principally in the achievement of a uniform product, i.e. in keeping constant the weight of the glass portions, this being dependent on many different factors. That these difficulties have been overcome is demonstrated by the fact that the duties of the turntable-man are limited to the occasional adjustment of a tap, checking the weight of a sample bulb, etc.

THE ROLE OF THE ELECTRON MICROSCOPE IN THE INVESTIGATION OF PLANT DISEASES

by E. van SLOGTEREN *).

632.3:581.23:621.385.833

To emphasize the usefulness of the electron microscope in the realm of biology, would be carrying coals to Newcastle. Nevertheless, it may well be worth while to consider the role this modern tool of research is playing in a specialized field of biology, phytopathology. The Laboratory for Flowerbulb Research at Lisse, which under the direction of Prof. Dr. E. van Slogteren has for a long time held a prominent place in this field, is one of the institutes availing themselves of the new possibilities offered by electron microscopy. The following invited paper, while providing the readers with a very succinct picture of phytopathological research, will chiefly illustrate the potentialities of the electron microscope by presenting a number of excellent electron micrographs obtained by the workers at Lisse.

Among the numerous varieties of tulips occurring in the standard collections of our florists a number of varieties of the so-called "broken" type are encountered. These are characterized by the petals containing differently coloured parts, which may exhibit various combinations of colours as well as various patterns (flamed, streaked, pinnated, speckled, spotted, etc.). In *fig. 1* a beautiful broken tulip, of the variety Cottage Tulp Perseus, is shown.

These types of tulips are by no means new. The earliest known description of broken tulips was given in 1576 (Clusius). When, in the early seventeenth century, tulips came into vogue, especially in France and in the Netherlands, broken tulips had their share in it. This is testified inter alia by the flower paintings made in those years, as for instance the numerous pictures by the members of the Bosschaert family (father and three sons); one of these pictures is illustrated in *fig. 2*. Broken tulips even were the chief object of the famous tulipomania of the years 1634 to 1637.



Fig. 1. Broken tulip of the variety Cottage Tulp Perseus. (From: E. van Slogteren and Maria P. de Bruyn Ouboter, Mededelingen Landbouwhogeschool Wageningen 45, No. 4, 1941.)

*) Laboratory for Flowerbulb Research, Lisse (Netherlands).

Comparatively new, however, is the knowledge that broken tulips in fact are no varieties of their own, but specimens of common varieties, befallen by a disease, one of whose symptoms is the beautiful spotting or flaming of the petals.

Although in the broken tulip itself the disease apparently does not do much harm and in particular does not seem to affect greatly its reproductive power, the same will in general not apply to other crops (or even other varieties of tulips) to which the disease may be transferred in various ways, e.g. by aphides.

This statement should be amplified. The broken tulip is one of the rare though spectacular cases in which a plant disease may be economically useful. In the vast majority of cases, evidently the contrary is true. In fact, the importance of plant diseases as a virtual threat to human subsistence can scarcely be overestimated. Practically every crop is liable to infection not by one but by a variety of diseases, and the investigation of such diseases may safely be labelled as an objective of major importance of our scientific efforts. Many practical measures have already sprung from such an investigation, ranging from the precautions imposed on the cultivating and marketing of broken tulips to the routine testing of seed-potatoes as to the presence of inter alia the potato X-virus. This test is now compulsory in the Netherlands for all seed-potatoes selected for export.

The potato X-virus: it will be known that many infectious plant diseases may be ascribed to viruses. Viruses are now known to be responsible for the breaking of tulips. Other diseases are caused by fungi, or by bacteria, as e.g. the Yellow disease of hyacinths (caused by *Xanthomonas hyacinthi* (Wakker) Dowson).

A virus is a parasitical agent found in living tissues and capable of reproduction and of transfer to other tissues. In these respects viruses are similar to bacteria. One of the main characteristics of a virus, however, distinguishing it from bacteria, is its extreme smallness. It was found that virus particles pass through all usual bacterial filters — a fact which in those years when the notion of a virus was only beginning to crystallize¹⁾ provided a means for concentrating the "agent" and which initially was thought to constitute one of its essential properties. At present some non filter-passing viruses are known, but most viruses are extremely small indeed, too small for observation

under the most powerful optical microscope. One of the smallest, the virus of the foot-and-mouth disease, is now known to consist of particles of size about 10 millimicrons.

To-day, more differences existing between viruses and bacteria are known. The most fundamental one is the fact that viruses are capable of reproduction only in living cells, while bacteria form colonies on dead matter, e.g. in broth. This gives evidence that bacteria must be living organisms, whereas for viruses this is still a question open to doubt: by many investigators the opinion is upheld that a virus is not an organism but must be considered simply as a gigantic protein molecule.

By the very smallness of the viruses (from a macroscopic point of view!) frustrating all efforts to make them visible under the microscope, research in this field for many years was curiously handicapped: the effects of the "parasitical agent" could clearly be traced, it could be transferred to other plants, its reaction to chemicals or to other "agents" could be studied — but nobody had ever set eyes on it. Diagnosis, which in the practical fight against plant diseases is all-important, had chiefly to be based on the study of symptoms. In later years it was in some cases greatly aided by the use of a serological reaction, that was developed at (among other laboratories) the Laboratory for Flowerbulb Research at Lisse²⁾. This reaction is based on the fact that plant viruses on being brought into animal blood, e.g. of rabbits, provoke the formation of specific antibodies. The serum containing these antibodies (antiserum), when added to a sample of plant juice containing the original virus, will cause an agglutination or precipitation, visible to the eye, of virus particles or plastids that have adsorbed the virus at their surfaces. Only in a limited number of cases, however, antisera could hitherto be obtained, while, on the other hand, even with a positive serological reaction, the diagnosis should not be considered complete as long as, in accordance with one of Robert Koch's postulates, the parasitical agent has not actually been shown in the host.

In these circumstances it is not surprising that, when the electron microscope became of age (and even earlier than that), phytopathologists eagerly seized the possibility offered by this instrument of discerning particles of submicroscopical size, down to 100 m μ or less. As early as in 1939,

¹⁾ D. Iwanowski, Ueber die Mosaikkrankheit der Tabakspflanze, Bull. Acad. Imp. Sci. St. Petersburg. N.S. III, 35, 67-70, 1892.

²⁾ E. van Slogteren, De betekenis van de serologie voor het virusonderzoek, T. Plantenziekten 49, 1-21, 1943 (in Dutch).



Fig. 2. Painting made by Ambrosius Bosschaert Jr in 1635. Among the flowers two broken tulips are to be seen. (Courtesy of the Director of the Centraal Museum, Utrecht. The reproduction was made by N.V. Wereldbibliotheek, Amsterdam.)

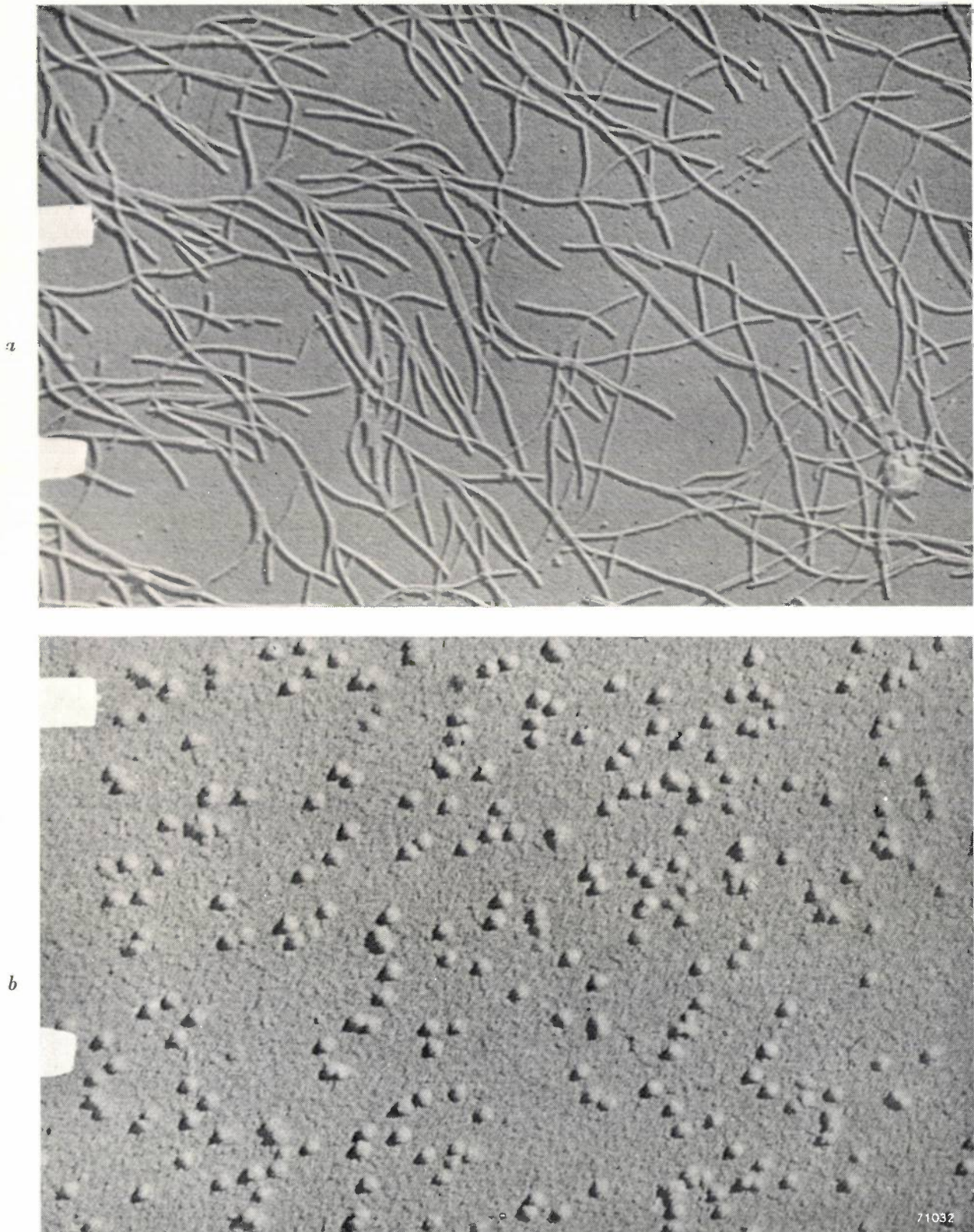


Fig. 3. a) Electron micrograph of the potato X-virus. Enlargement 30 000 \times . b) Carnation Yellow mosaic virus. Enlargement 50 000 \times . (This photograph and those of figs. 4, 5 and 6 were published in the article quoted in footnote 4).)

All electron micrographs reproduced in this paper were made with the Philips 100 kV electron microscope (cf. Philips techn. Rev. 12, 33-51, 1950). The samples of plant juice were put on a thin film of „Formvar” (polyvinylformal dissolved in $\frac{1}{4}\%$ solution of pure chloroform) covering the tiny slit of the object holders. All the samples shown were shadowcast with palladium. The distance between the blocks at the edge is in all cases 1 μ .

Kausche, Pfankuch and Ruska³⁾, using the electron microscope succeeded in rendering plant virus "particles" visible to the eye.

In recent years the electron microscope has been used in many laboratories for the investigation of plant diseases, and it has revealed quite a number of particulars concerning many kinds of viruses. The shape and size of virus particles has been studied (figs. 3a and b) the character of the serological reaction already mentioned has been investigated, the "crystallization" of some types of virus (e.g. the "bushy stunt" virus of tomato) has been clarified, etc.

Evidently the electron microscope was predestined to assume an important role in diagnostics. By this means, in many cases a contamination could conclusively be proved to be present in suspected plants and the virus involved could be identified⁴⁾. A typical case, encountered in an examination of infected tobacco plants, may be described as follows. Among the different known tobacco diseases, there are two of special importance: tobacco necrosis, associated with a virus consisting of round particles (fig. 4a), and the famous tobacco mosaic disease⁵⁾, caused by a rod-shaped virus (fig. 5a). Some of the tobacco plants under investigation gave a positive serological reaction with the antiserum of tobacco necrosis and not with the specific tobacco mosaic antiserum; these plants showed in their juice the round virus particles of tobacco necrosis. Other plants reacted positively with tobacco mosaic antiserum and not with tobacco necrosis antiserum; in the juice of these plants the rod-shaped particles of tobacco mosaic virus were found. Still other plants, while not exhibiting the typical symptoms of the disease, reacted positively with both antisera, and indeed an electronmicrograph of their juice revealed the presence of both the round and the rod-shaped virus particles.

The nature of the agglutination observed in these serological reactions is clearly shown by the micrographs fig. 4b and 5b.

The electron microscope proved quite useful also for recognizing and identifying bacteria causing plant diseases. These relatively large objects can

easily be made visible, the only difficulty being to get them in full dress in the picture, with their characteristic cilia (fig. 6). When from a preliminary investigation with the ordinary microscope the presence of a specific kind of bacteria is expected, it may well be possible with the electron microscope to identify these bacteria amidst the multitude of ever-present innocuous bacteria, thus saving time by avoiding the necessity of isolating and cultivating the expected bacteria.

Of course, the identification of the agent, especially in the case of a virus disease, will not always be so simple. Too often particles found in the juice of a diseased plant and suspected of being the virus causing the disease are found also in the juice of apparently healthy plants. Thus, if the particles belong to a parasitic agent, it must be an innocuous one, carried by both diseased and healthy plants. In that case some other, undetected, particle may be the cause of the disease. The possibility of the agent proper escaping detection can by no means be ruled out, as in quite a number of cases it has not been possible to find anything even faintly resembling a virus particle in the samples made of the juice of severely diseased plants. The real problem, then, resides in preparing the samples: in the course of the different treatments, such as ultracentrifugation or precipitation with different chemicals, applied to the plant juice in order to enrich its virus content by removing undesired components (plant protoplasm etc.), it may happen that the virus is lost, as for instance it may not easily part with the protoplasm. Moreover, the possibility should not be excluded that the different treatments might change the size and shape of the virus particles or might produce "artefacts" in the picture, which in comparing these particles with other types, known before, obviously may give rise to serious misinterpretations.

To conclude and summarize these remarks it should be stated that the power of the electron microscope can only be put to good use in plant virology when combined with all other methods of approaching the many-sided problems in this field. Experience of different laboratories in the preparation of samples should be exchanged, and no electron micrograph of particles supposed to belong to a new type of virus should be given without a short indication as to how the material has been treated. When such precautions will have eliminated all possible doubt concerning the identity between the original object and the picture obtained, the electron microscope, besides providing a valuable assistance in routine examinations,

³⁾ G. A. Kausche, E. Pfankuch and H. Ruska, Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop, Naturwiss. 27, 292-299, 1939.

⁴⁾ Maria P. de Bruyn Ouboter, J. J. Beijer and E. van Slogteren, Diagnosis of plant diseases by electron microscopy, Antonie van Leeuwenhoek 17, 189-208, March 1951.

⁵⁾ Cf. footnote¹⁾ and a great number of papers published by W. M. Stanley in the years between 1930 and 1940 in Science, J. biol. Chem. and other journals.

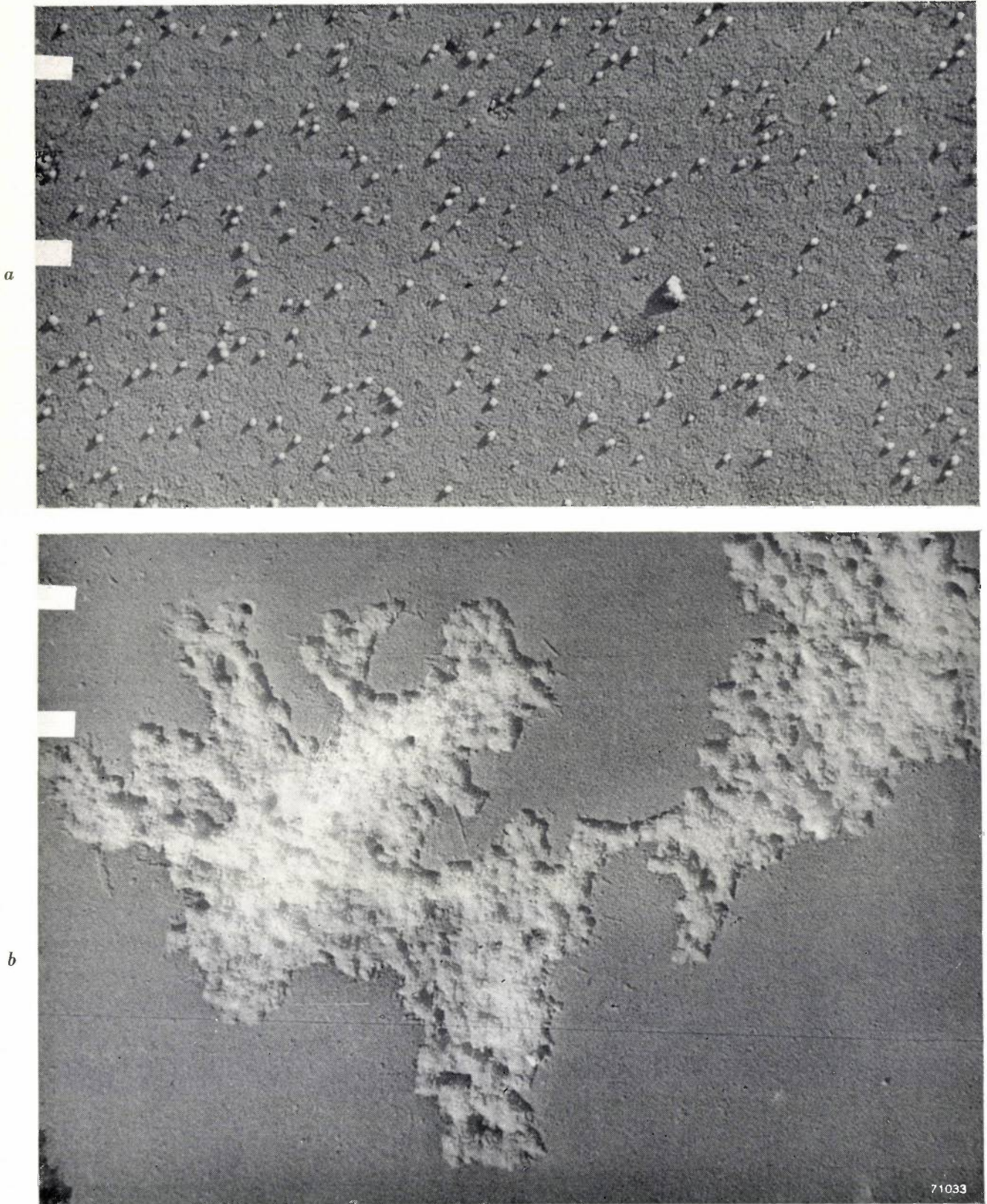


Fig. 4. *a*) Tobacco necrosis virus, obtained from tobacco plants (this virus is also known as causing the Augusta-ziek of tulips). Enlargement 28 000 \times . As an example, the complete procedure of purification is indicated in this case: the foliage were minced, frozen for a night, squeezed, the juice centrifuged for 30 minutes at 3500 rev/min, purified with ammonium sulphate, centrifuged for 30 min at 3500 rev/min, 5 \times concentrated by centrifuging for 40 min at 40 000 rev/min, and dialysed one night. *b*) Same virus as (*a*), showing positive serological reaction with tobacco necrosis antiserum. Enlargement 18 000 \times .

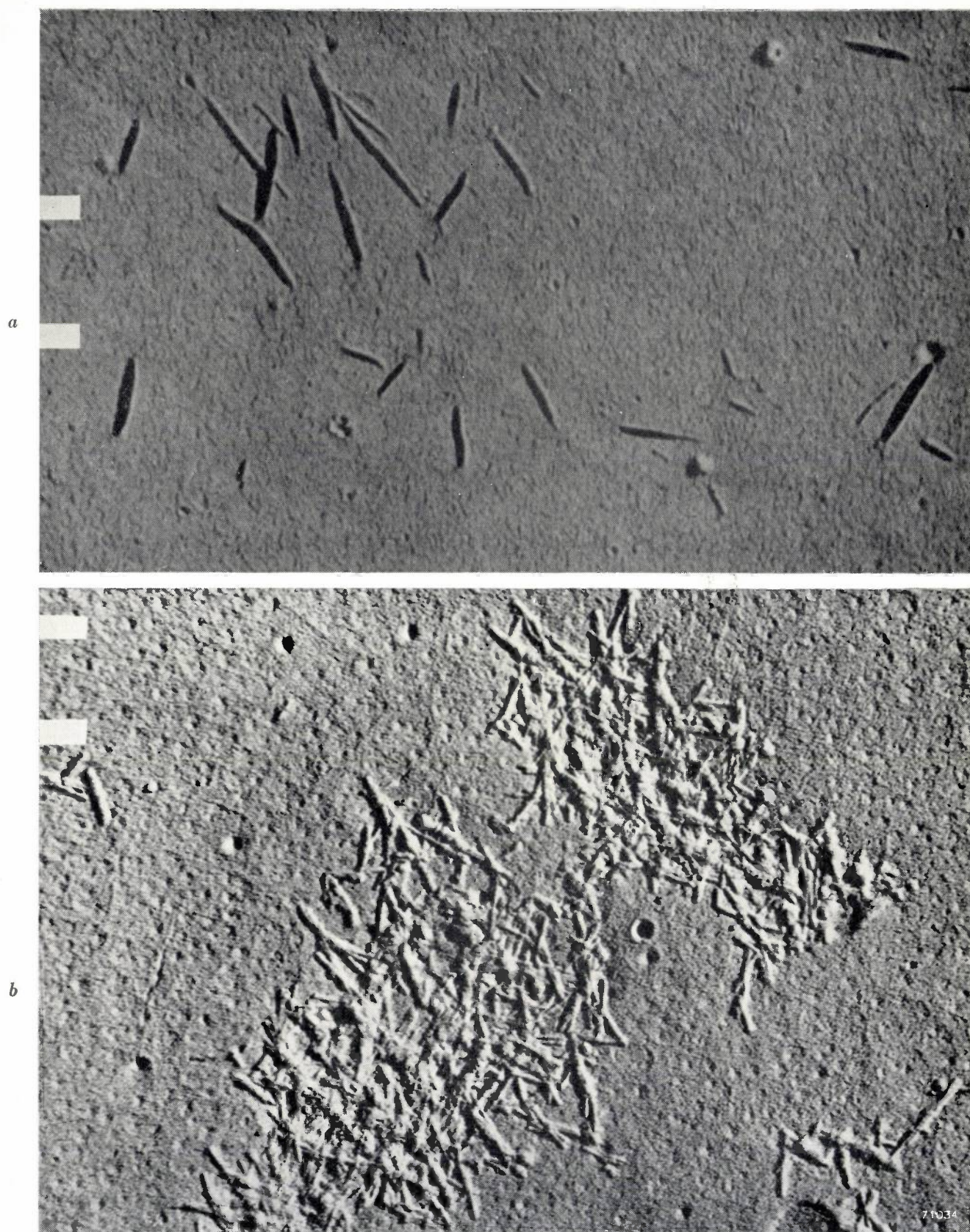


Fig. 5. *a*) Tobacco mosaic virus, obtained from tomatoes. To the solution obtained a quantity of normal serum (containing no antibodies) was added: negative serological reaction. Enlargement 18 000 \times .
b) Same virus as (*a*), showing positive serological reaction with tobacco mosaic antiserum. Enlargement 15 000 \times .

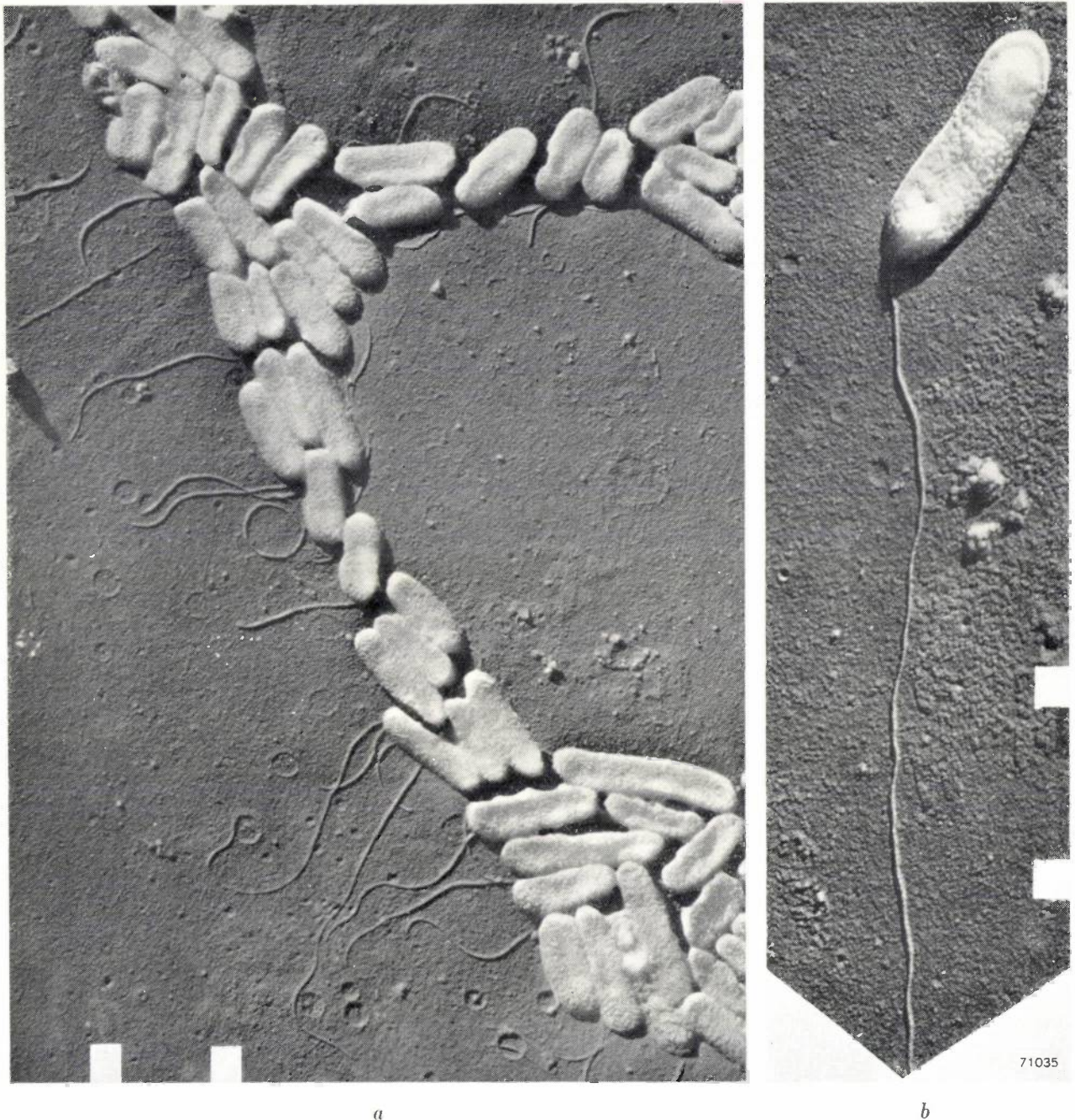


Fig. 6. *Xanthomonas hyacinthi* (Wakker) Dowson, the bacteria causing the Yellow disease of hyacinths. Plant tissue with bacteria in distilled water.

a) Enlargement 14 000 \times .

b) Other part of sample, enlargement 23 000 \times .

may well allow man to venture farther than ever down the fascinating road which at its end is bordered by withering crops and strangely coloured flowers, while its origin is covered by the nebulae of the problem of the nature of life itself.

Summary. Although a considerable amount of knowledge of plant viruses had already been gathered since their discovery in 1892, it was only in 1939 that for the first time virus particles

were made visible to the eye. This was achieved by using the electron microscope. Ever since then, electron microscopy has been applied in phytopathology and important results have been obtained. Problems investigated with the electron microscope were, for instance: the shape and size of different kinds of viruses, the reaction of viruses with their specific antisera, the apparent crystalline form of some viruses. In this paper stress is laid on the usefulness of the electron microscope for the diagnosis of plant diseases, while at the same time essential conditions and limitations in this application are put forth. The paper is illustrated by a number of electron micrographs obtained at the Laboratory for Flower-bulb Research at Lisse with the aid of the Philips 100 kV electron microscope.

HIGH-TENSION GENERATORS FOR LARGE-PICTURE PROJECTION TELEVISION

by J. J. P. VALETON. 621.397.62:621.314.5:621.396.615.12.072.2

For the projection of television pictures which are not only large but also of high definition and brightness the cathode-ray tube requires a high voltage, of 25 to 50 kV. Generators for supplying such high voltages must conform to a number of very special requirements which have led to the development of other methods of producing such voltages than the more obvious system of stepping up and rectifying the mains supply. A few methods of achieving this purpose are outlined and one of these is described in detail.

The potentials employed for the acceleration of the electrons in television tubes have in the course of time become higher and higher. Whereas 5 kV was normal for a direct-vision tube in 1945, this was soon increased to 7 or 9 kV. Nowadays 12 kV or 14 kV is more usual and there are indications that this might be increased to even 16 or 17 kV.

This tendency towards higher potentials is the outcome of the constant demand for increased brightness, definition and dimensions in television pictures; to produce a brighter picture the cathode ray must deliver more energy to the fluorescent coating on the screen per unit of time, so that either the beam current or the acceleration voltage, or both, must be increased. A higher beam current, however, has an adverse effect on the definition, whereas a higher potential facilitates focusing, a factor which is particularly important when it is a question of tubes of the shortest possible length with a large screen. In such cases the deflection angles are wider and, in order to ensure good definition at the edges, a very slender beam must be employed; this is made possible to an increasing degree by reducing the current and increasing the acceleration voltage.

The use of a tube with a large screen is not the only means of obtaining large pictures, however. A very bright, clear picture can be produced in a tube of quite small size, and this can be reproduced on a separate screen by means of a highly efficient optical system. This is the method employed in a projection television system developed by Philips, of which a description has already appeared in this Review¹⁾. To be able to meet the high requirements imposed on it as regards brightness and definition, the tube in question, the MW 6-2²⁾, has to operate

with an acceleration voltage of 25 kV, the average current being at most 200 μ A. The brightness and definition of the picture on the screen, which is 6 cm in diameter, are sufficient to admit of optical enlargement to give a projected picture having a diagonal of 50 cm.

In recent years the Philips Laboratories at Eindhoven have evolved a projector along these lines which is capable of giving a picture measuring 3 m by 4 m. The tube employed in this equipment has a screen 12 cm in diameter and the current on which it operates is considerably higher than that of the MW 6-2; an acceleration voltage of 50 kV was found necessary to secure adequate focusing of the beam.

The above introductory remarks sufficiently illustrate the reasons for the upward trend of acceleration voltages. On the other hand, however, certain inhibiting factors are encountered; according as the velocity of the electrons is increased, more power is required to deflect the beam through a given angle, and this means more expensive deflection equipment. Then again, when the acceleration potential is raised it becomes increasingly difficult to manufacture tubes that will withstand such potentials and to devise generators to supply the required voltage. Such generators form the subject of the present article.

Methods of generating high tension

Let us now review the various possibilities of developing the high potentials required for television receivers.

Rectification of A.C. voltage at mains frequency

In early television receivers the required direct voltage of 5 kV was obtained by rectifying a high alternating voltage taken from a transformer operating on the mains supply, but this has numerous disadvantages.

To secure from the low-frequency mains a rectified

¹⁾ Projection-television receiver, I. The optical system for the projection, by P. M. van Alphen and H. Rinia, Philips techn. Review 10, 69-78, 1948.

²⁾ Projection-television receiver, II. The cathode-ray tube, by J. de Gier, Philips techn. Review 10, 97-104, 1948.

voltage with a sufficiently small ripple, high-value smoothing capacitors are essential and these are not only costly components, but their large store of energy may prove to be a cause of damage in the event of a breakdown in the circuit. Another drawback of the low frequency is the large number of turns required for the power transformer, made necessary by the low value of the induced voltage per turn. The heater winding for the rectifier valve, further, necessitates high-tension insulation and this, together with the H.T. winding itself, takes up a great deal of space.

Almost universally, therefore, this method of producing the required voltage was discarded as it became apparent that the high tension could be produced fairly simply as a "by-product" of the equipment serving for the magnetic deflection of the electron beam.

High tension as by-product of the deflection system

Deflection of the electron beam, both vertically and horizontally, is usually effected by means of magnetic fields generated by saw-tooth currents flowing in the deflection coils. The rapid flyback of the luminous spot from the right-hand side of the screen (end of line) to the left-hand side (start of next line) is the result of a sudden jump in the current flowing in the line-deflection coil. This simultaneously produces a voltage peak across the coil in question which may be more than 10 times higher in value than the voltage across the coil during the scanning of the line. To match the coil to the generator of the saw-tooth current a step-down transformer is connected between the two. On the primary side the peaks reach as much as 5 kV and, by extending this winding in the form of an auto-transformer, it is a simple matter to obtain still higher direct voltages after rectification, say roughly 8 kV. A rectifier circuit for voltage-doubling (cascade) then offers the possibilities of obtaining voltages up to 10 or even 14 kV.

The great advantage of this arrangement is to be found in the high frequency at which the voltage peaks occur in the horizontal deflection; with 625 (405) lines and 25 frames per second, for instance, the repetition frequency will be $625 \times 25 = 15,625$ c/s ($405 \times 25 = 10,125$ c/s). Small capacitors will therefore effectively smooth the rectified voltage³⁾ and such capacitors are not only inexpensive but, because of their relatively low charge, they involve no hazards.

³⁾ Moreover, the ripple voltage (if not too high) interferes very little with the picture since it is synchronous with the horizontal motion of the spot.

This method is particularly suitable when the power required for the electron acceleration is small compared with the reactive power used for deflection purposes, that is, so long as the first can be justifiably qualified as being a by-product of the deflection system. With direct-vision tubes this is indeed the case and where these tubes are used the method in question is therefore almost universally employed.

Rectification of separately generated pulses

The problem as applicable to tubes for projection television is rather different. Here the power required for acceleration is no longer small compared with the reactive power required for the deflection. Should it nevertheless be required to adopt the method described, the deflection equipment will have to be larger than would otherwise be necessary. Furthermore, care must be taken that fluctuations in the average beam current — which corresponds to the average brightness of the picture — cannot cause the high tension or the deflection amplitude to vary. For these reasons preference is usually given to other systems.

One such system, suitable for domestic projection television has already been described in detail in this Review⁴⁾. In this particular instance the direct voltage is 25 kV, this being obtained from three stages in cascade. Pulses having a peak value of about 8.5 kV are fed to the first stage, these being obtained by passing a saw-tooth current through a coil. With each current pulse in the coil, which with its self-capacitance constitutes an oscillatory circuit, a damped oscillation is set up, both half cycles of which are rectified. The resonant frequency of the circuit is about 32,000 c/s, the repetition frequency with which the circuit is excited being roughly 1000 c/s.

For further details the reader is referred to the article mentioned in note 4).

Rectification of a high sinusoidal voltage of high frequency

A fourth method consists in rectifying a high sinusoidal alternating voltage as provided by an oscillator (with or without cascade circuit). This arrangement is also sometimes met with in projection television receivers. Since the coils used are generally of the air-cored type, there is no alternative but to operate the oscillator at a frequency which is higher than required for effective smoothing, for example 300 kc/s, the size of the coils being

⁴⁾ Projection-television receiver, III. The 25 kV anode voltage supply unit, by G. J. Siezen and F. Kerkhof, Philips techn. Review 10, 125-134, 1948.

nevertheless such that very close coupling cannot be secured; the efficiency of the oscillator is therefore fairly low and the internal resistance (on the D.C. side) high. As we shall see presently, however, a low internal resistance is essential; the higher the inherent resistance proves to be, the greater the amount of trouble involved in reducing it.

Considerable improvement can be obtained in this respect by employing Ferroxcube magnetic materials⁵⁾ for the oscillator coils. In this way it is possible to operate with relatively low frequencies (some tens of kc/s) without the need for coils of large dimensions⁶⁾, at the same time ensuring low magnetic losses in the core and low dielectric losses in the insulation. From the point of view of smoothing, there is no object in employing a higher frequency, since a certain amount of reserve capacitance is in any case essential to meet the fluctuations in the load.

In the following this method of producing high tension is discussed in greater detail, this being followed by a description of two existing designs for the supply of 50 kV high tension. First, however, let us review the electrical requirements to be imposed on H.T. generators for television receivers.

Electrical requirements to be met by a T.V. high-tension generator

In the first place the generator must naturally be capable of delivering the required amount of current at the voltage desired. As far as the current is concerned, it is usual to base this on the highest mean value of the beam current for which the tube is rated; the peaks may be several times higher. It is often stipulated that the generator must be able to supply this mean current ($I_{0 \max}$) at the nominal value of the acceleration voltage ($V_{0 \text{ nom}}$); further that the voltage shall drop rapidly in the event of the current assuming a constant higher level (so that the short-circuit current will not largely exceed $I_{0 \max}$), and that the no-load voltage shall be little greater than $V_{0 \text{ nom}}$. Fig. 1 illustrates the required relation between the voltage V_0 and the current I_0 . A characteristic of this kind provides as it were a safety precaution against overloading of the screen of the cathode-ray tube.

For so far as the instantaneous value of the current, relating to the high lights of the picture,

exceeds the average, the current is supplied by a reservoir capacitor. Sometimes the usual capacitors in the rectifier circuit will provide sufficient reserve capacitance, or the anode capacitance of the tube itself may fulfil this function, but, if not, extra capacitance must be provided, as will be shown presently.

In the second place, the internal resistance is very important. It has already been mentioned that this resistance should be low, i.e. that the characteristic (fig. 1) should be roughly horizontal from $I_0 = 0$ to $I_0 = I_{0 \max}$. The reason for this is two-fold.

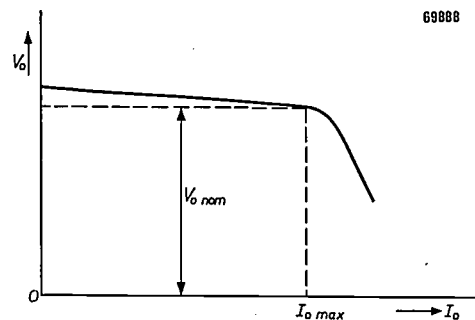


Fig. 1. Desired V_0 - I_0 characteristic for an H.T. generator for television receivers. The no-load voltage should be but little more than the voltage $V_{0 \text{ nom}}$ relating to the highest average current $I_{0 \max}$ of the tube, and should drop rapidly when more current than that is taken.

In television tubes the electron beam is generally focused by means of a magnetic "lens". The power of this lens is dependent on the acceleration voltage; if this voltage varies, for example as a result of a change in the average brightness of the picture, adjustment must also be effected to keep the definition as high as possible. Hence, if the internal resistance is too high, the focus must be constantly readjusted.

The other consideration concerned is that, in the magnetic deflection employed with most tubes today, the deflection angle is also dependent on the acceleration voltage. (This angle is inversely proportional to the square root of the acceleration voltage). If this voltage varies, the picture size will fluctuate and this is quite as unpleasant as a varying focus.

To obviate these defects care should be taken to ensure that the acceleration voltage varies as little as possible with the current or, in other words, that the internal resistance of the generator shall be as low as possible.

Some further aspects of an H.T. generator comprising an oscillator and a rectifier

Let us now look a little more closely at the last

⁵⁾ J. L. Snoek, Non-metallic magnetic material for high frequencies, Philips Techn. Review 3, 353-360, 1946, and W. Six, Some applications of Ferroxcube, Philips techn. Review 13, 301-311, 1952 (No. 11).

⁶⁾ The possibilities in this direction were indicated by P. J. H. Janssen of the Television Development Laboratory, Eindhoven.

mentioned method of producing an H.T. voltage, taking first the rectifier, then the oscillator and, finally, measures to ensure low internal resistance.

The rectifier

As starting point we will take the simplified circuit depicted in fig. 2a, in which $L-C$ is the oscillatory circuit, D is a diode and C_1 is a smoothing capacitor. This oscillatory circuit is regarded at the same time as part of the rectifier circuit, because a conclusion will be drawn about the values of L and C from the following considerations.

The time t is taken to be zero at the instant when the voltage v across the oscillatory circuit is zero and on the verge of becoming positive (i.e. when the side of C that is connected to the anode becomes positive with respect to the other side). The voltage v is thus represented in the first instance by: $v = V \sin \omega t$ (fig. 2b) and, disregarding the losses in the circuit, the current i will be: $i = I \cos \omega t$ (fig. 2c). In this, V is the peak value which the voltage would attain in the absence of the rectifier; I denotes the peak value of the current and ω the natural angular frequency of the oscillatory circuit.

As soon as v becomes equal to the potential across the smoothing capacitor, the diode passes current. For the sake of convenience we will regard the capacitor C_1 as being so large that the current pulses from the diode do not substantially alter the voltage across C_1 ; in other words, that a direct voltage (V_0), the ripple of which may be disregarded, occurs across C_1 . We shall also disregard the voltage drop across the diode when current is passing. As long as current flows in the diode, the voltage across C and L will therefore be the direct voltage V_0 .

The fact that the voltage across C is constant means that no current passes through this capacitor, that is, that the current i flowing in L passes in its entirety through the diode. The constant voltage across L implies that this current i (which is equal to the diode current i_d , fig. 2d) decreases linearly with time.

The phase angle ωt at which the diode begins to pass current will be denoted by $\pi/2 - \varphi$; then:

$$V_0 = V \cos \varphi, \dots \dots \dots (1a)$$

the initial value I_d of the diode current being

$$I_d = I \sin \varphi \dots \dots \dots (1b)$$

When $t = 0$, only magnetic energy exists in the oscillatory circuit, namely: $\frac{1}{2} LI^2$. A part of this, viz. $\frac{1}{2} CV_0^2$, appears as electric energy in the capacitor C , whilst the remainder, $\frac{1}{2} LI_d^2$, is applied

to the smoothing capacitor by way of the diode. Hence:

$$\frac{1}{2} LI^2 = \frac{1}{2} CV_0^2 + \frac{1}{2} LI_d^2.$$

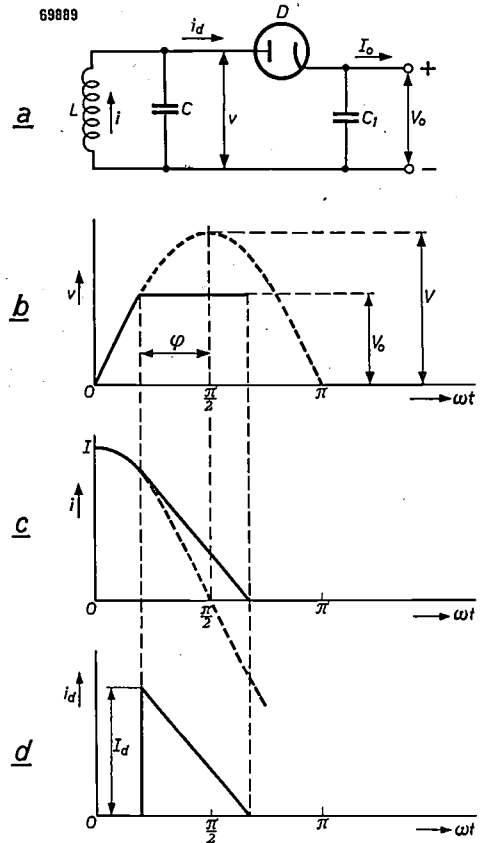


Fig. 2. a) Circuit of an H.T. generator. $L-C$ = oscillatory circuit forming also part of the oscillator. D = diode; C_1 = smoothing capacitor.

The following quantities are shown as functions of ωt : b) the voltage v across the $L-C$ circuit; c) the current i flowing in inductance L ; d) the current i_d passing through the diode. The curves are not continued beyond the point where i_d becomes zero.

In conjunction with (1b) this gives:

$$I_d^2 = \tan^2 \varphi \cdot \frac{C}{L} V_0^2. \dots \dots \dots (2)$$

Let f denote the frequency with which the cycle is repeated; the capacitor C_1 then receives f times the energy $\frac{1}{2} LI_d^2$ per second and, in the same interval, the energy $I_0 V_0$ is withdrawn from it (I_0 is the mean value of the direct current). In the stationary condition, then:

$$f \cdot \frac{1}{2} LI_d^2 = I_0 V_0 \dots \dots \dots (3)$$

If φ is small, i.e. as long as C_1 is in parallel with C only for a small fraction of the cycle, we may say as an approximation that

$$f = \omega/2\pi = 1/(2\pi\sqrt{LC}).$$

From (3) it then follows that:

$$I_d^2 = 4\pi \sqrt{\frac{C}{L}} I_0 V_0 \dots \dots (4)$$

Elimination of V_0 from (4) and (2) gives:

$$\tan \varphi = 4\pi \frac{I_0}{I_d}, \dots \dots (5)$$

and subsequent elimination of I_d^2 gives:

$$\sqrt{\frac{L}{C}} = \frac{\tan^2 \varphi}{4\pi} \cdot \frac{V_0}{I_0} \dots \dots (6)$$

These equations will serve as a guide in the design. For I_0 and V_0 we take the highest mean current $I_{0\max}$ and acceleration voltage $V_{0\text{nom}}$ required for the cathode-ray tube; we first decide upon $\tan \varphi_{\max}$, that is, the value of $\tan \varphi$ which corresponds to $I_0 = I_{0\max}$, taking into consideration the following.

Under the no-load conditions the direct voltage is equal to the peak value V of the alternating voltage, whilst on full load it is equal to $V \cos \varphi$. Now, in order to keep the internal resistance of the generator as low as possible without resorting to special measures, φ_{\max} should be made as small as possible, for the difference between V and $V \cos \varphi_{\max}$ will then be at a minimum. A lower limit for φ_{\max} may be derived from (5) by taking for I_d the maximum permissible peak current for the particular diode used; at lower values of φ_{\max} the diode would be overloaded.

Another consideration, however, prompts us not to take φ_{\max} to the lower limit; the efficiency is impaired when φ_{\max} is reduced, as will be seen from the following. In contrast with what has been assumed, the oscillatory circuit is in actual fact subject to certain losses. For a given (finite) quality, a certain part of the energy $\frac{1}{2} LI^2$, which is available when $t = 0$, is lost every cycle. The smaller the value of φ_{\max} , the smaller also the fraction of $\frac{1}{2} LI^2$ delivered per cycle by the diode, and the more unfavourable the ratio of this fraction to the losses. From the point of view of efficiency, the indicated procedure is therefore to employ a high value of φ_{\max} .

A compromise must therefore be sought, and a suitable value for $\tan \varphi_{\max}$ will usually lie somewhere between 0.2 and 0.5.

When $\tan \varphi_{\max}$ has been fixed, equation (6) gives the value of $\sqrt{L/C}$. By fixing a value for f as well, \sqrt{LC} is determined and hence L and C are known individually. With regard to the choice of f , the following should be noted.

In the foregoing it has already been said that a

high frequency is favourable from the point of view of smoothing. L and C , or at any rate LC , will then be small. But a disadvantage of a high frequency is that the losses in coils usually increase with frequency. By way of example, fig. 3

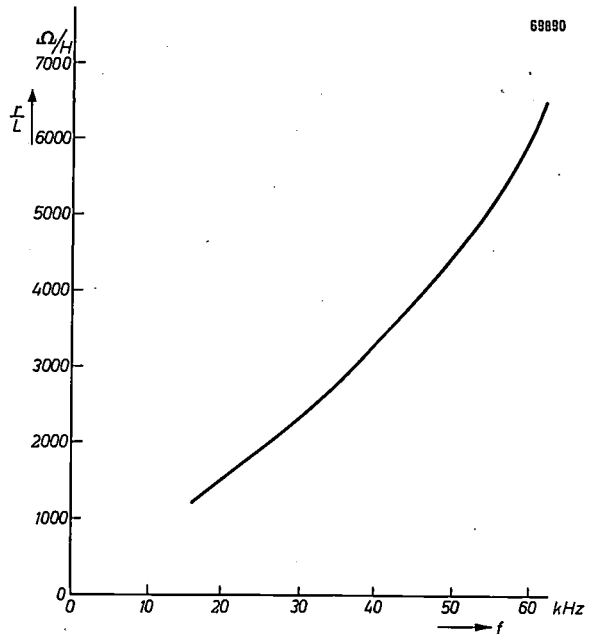


Fig. 3. Ratio of the loss resistance r to the inductance L as a function of the frequency f in a disc type of coil with Ferroxcube core.

illustrates the measured ratio of the loss resistance r to the inductance L of a disc type of coil, as a function of frequency. In order to keep the losses within bounds, then, the frequency should not be too high, but should preferably be above the highest audible frequency, so that whistles may be avoided in the possible event of mechanical vibration of any of the components.

The value of C as derived from $\sqrt{L/C}$ and \sqrt{LC} should of course not be less than the sum of the self-capacitance of the coil and the input capacitance of the rectifier. The last-mentioned contribution may be appreciable in the case of a cascade circuit using a large number of diodes. In the H.T. generators constructed in the laboratory we have endeavoured to arrange matters so that the input capacitance of the rectifier and the self-capacitance of the coil together just equal the computed value of C , thus dispensing with the need for an extra capacitor. This is worth doing, seeing that a capacitor suitable for the high voltages and high frequency in question is a costly item⁷⁾.

⁷⁾ It would of course be possible to connect a low-voltage capacitor of higher value across a part of the coil, but, owing the fact that the coupling between the coil sections is never perfect, this might give rise to parasitic oscillations. In view of this such a solution is better avoided.

The preferred frequency is between 20 and 30 kc/s, in which range the use of Ferroxcube for the cores has many advantages⁸). Compared with air-cored coils, fewer turns of wire are required in order to secure the desired inductance, whilst coupling between sections of the coil mutually and between the coil and other windings (which we shall discuss presently) is much closer. The eddy currents in Ferroxcube are negligible, moreover, so that in the frequency range concerned it is only necessary to take into consideration the hysteresis losses and dielectric losses in the insulation. In view of the hysteresis losses in particular, the flux density should not be too high, say, 0.10 to 0.15 Wb/m². The insulating material should have only a small loss angle.

So much then for the circuit employing a single stage of rectification (fig. 2a). The required direct voltage may, however, be so high that equation (6) will yield an impracticably high value of L/C , in which case the solution lies in the use of a cascade circuit, in which the total direct voltage is a multiple of the voltage V_0 in equation (6).

For details of the cascade arrangement the reader is referred to the article mentioned in note⁴), having regard to the fact that in the instance in question the alternating voltage applied to the input is a damped oscillation; in the present case we are concerned with an undamped oscillation.

The oscillator

In our discussion of the kind of rectifier depicted in fig. 2a, we have assumed that, at a time $t = 0$, a current I flows in the coil, without considering the manner in which this current originates. In the half-cycle during which the voltage v is positive — and if a cascade circuit is used, in the other half-cycle as well — the oscillatory circuit delivers energy

to the rectifier, but energy is also lost through the losses in the network. If this were not replaced, the current, after one cycle, would be less than the original value I .

It is the function of the oscillator to supply to the circuit just so much energy per cycle as is delivered and lost, so that the current in the coil will return to the value I at the commencement of every cycle.

A simple oscillator circuit is depicted in fig. 4. Both the anode and control grid circuits of a pentode are coupled inductively to the $L-C$ circuit in such a way as to ensure positive feedback. The circuit operates on the class C principle, that is, anode current flows for less than half a cycle. During the time that current flows, the anode voltage V_a is lower than the supply voltage V_B (fig. 5) and energy is then supplied to the circuit.

To improve efficiency it is useful to limit the passage of current to a mere fraction of a half-cycle.

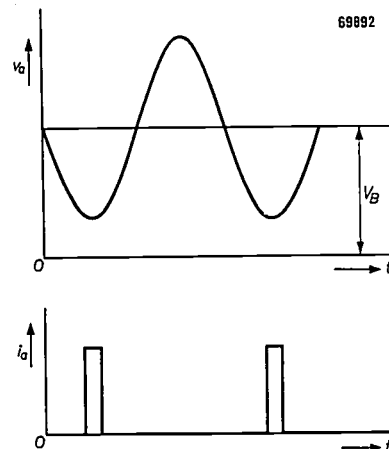


Fig. 5. Anode voltage v_a and pulsatory anode current i_a of the pentode P (fig. 4) as functions of time t (class C operation). The pulse i_a as shown in the diagram is exaggeratedly narrow.

That this does favour the efficiency may be seen from the following. Suppose that the anode-current pulses are so narrow that during these periods the current as well as the voltage may be regarded as constant, as in fig. 5. It is then a question of the ratio of the energy delivered to the oscillatory circuit per cycle, to the energy dissipated at the anode in the same period. From the above premise it follows that this ratio is the same as the ratio of the voltages v_{L_a} (across the coupling coil L_a) to v_a , both as regarded at the instant of the pulse. This ratio and accordingly also the efficiency is greatest when v_{L_a} is at its maximum, and v_a at a minimum, that is to say, when the anode current pulse occurs at the time when v_a is lowest. If the pulse be made any wider, so that v_a and v_{L_a} can no longer be regarded as constant during the period of the pulse, v_{L_a} and v_a in the above-mentioned energy ratio must be replaced by an average in respect of the duration of the pulse. This average will be less than the maximum for v_{L_a} and above the minimum for v_a , so that the ratio is then not so favourable.

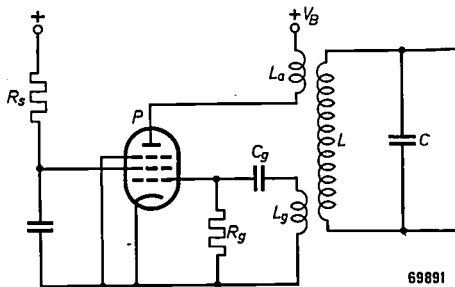


Fig. 4. Oscillator circuit. P = pentode with anode and control grid circuits coupled by means of the coils L_a and L_g to the oscillator circuit $L-C$ (cf. fig. 2a). C_g = grid capacitor. R_g = grid leak. R_s = resistor in series with the screen grid.

⁸) See article by W. Six referred to in note⁵).

Class C operating conditions are ensured by seeing that the alternating voltage across the coil L_g is several times greater than the grid base of the pentode.

A variable load on the H.T. generator will mean variable damping in the oscillator circuit or, generally speaking, variations in the amplitude of the alternating voltage, manifested as a fluctuating direct voltage. Apart from special measures to ensure a constant direct voltage, as discussed in the next section, precautions can be taken in the design of the oscillator itself to make the alternating voltage as independent as possible of the damping by suitably limiting the amplitude.

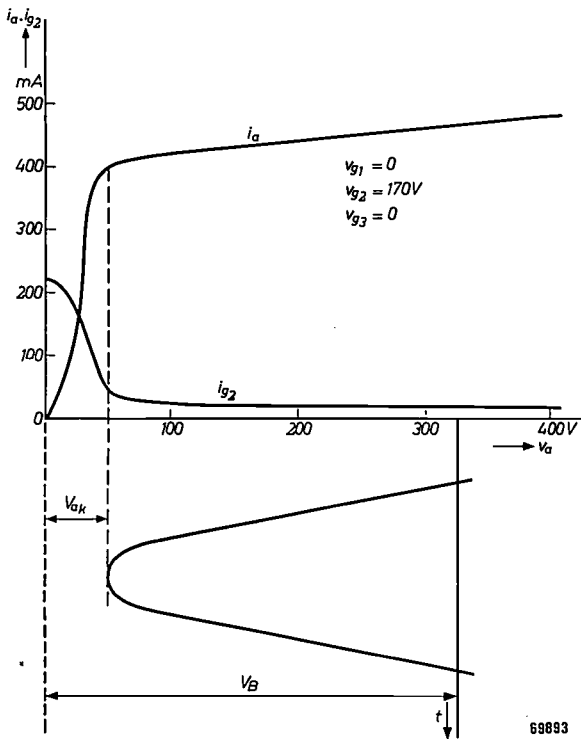


Fig. 6. Anode current i_a and screen grid current i_{g2} of the pentode PL 81 as functions of the anode voltage v_a , with constant screen voltage $V_{g2} = 170$ V and zero volts on control and suppressor grids.

In the oscillator, v_a varies with the time t sinusoidally about the value V_B . The amplitude as shown is so large that the minimum of v_a lies just at the bend in the i_a - v_a characteristic (V_{ak}). In this way the amplitude of v_a is limited.

When an oscillator is switched on, the voltage amplitude rapidly increases up to the point where it is limited by one factor or another. For our particular purpose this limitation can best take place by the bend in the anode current i_a versus anode voltage v_a characteristic of the pentode, the grid potentials being constant (fig. 6). Limitation is established by reason of the fact that when the amplitude of the alternating anode voltage in-

creases, the minimum of v_a ultimately becomes so small that it lies at the bend in the i_a - v_a curve. Any further increase in the amplitude is checked because v_a would enter the zone in which i_a drops rapidly with decreasing v_a .

The limiting effect can be increased taking advantages of fact that, below the bend in the curve, the screen-grid current i_{g2} rises sharply at the expense of i_a when v_a drops. If the screen grid is fed through a resistor (R_s , fig. 4), the screen voltage will drop when amplitudes greater than that shown in fig. 6 occur. In consequence, the i_a - v_a characteristic is displaced to a lower level and i_a decreases to a greater extent than if the screen voltage were constant.

If the coils L and L_a (fig. 4) are very closely coupled — and this is quite possible when Ferroxcube cores are used — the amplitude V of the alternating voltage across the L - C circuit, and therefore also the direct voltage, will be nearly enough proportional to the amplitude of the alternating anode voltage. In the event of small variations in the last-mentioned voltage, V_0 will also vary but slightly.

The circuit can best be so constituted that when the direct current withdrawn from it is at a maximum (that is, with maximum damping of the oscillatory circuit), the minimum value of v_a will lie just at the bend in the i_a - v_a curve. On smaller loads the amplitude of the alternating anode voltage will tend to increase, since, per cycle, more energy is then supplied to the oscillatory circuit than is being taken from it. This would result in an undesirable increase in the rectified high tension, but the limiting devices described above ensure that the slightest increase in the amplitude of the alternating anode voltage results in a decrease in anode current, sufficient to restore equilibrium between the energies received and delivered per cycle. The curve representing i_a as a function of time then reveals a dip which is so much the deeper according as the direct current is lower (fig. 7).

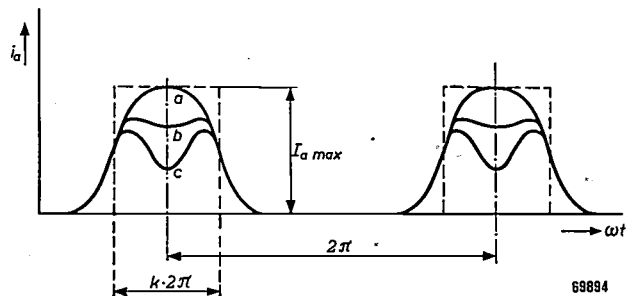


Fig. 7. Anode current i_a of the pentode referred to in fig. 4 as a function of time: a on full load, b on half load and c without load.

In the circuit under review the pentode used should preferably be one that will pass a heavy anode current at quite a low anode potential. At the same time, the bend in the i_a-v_a characteristic should be as sharp as possible and should occur at a low value of v_a . The pentode PL 81 satisfies these conditions.

In order to facilitate computation of the oscillator circuit we replace the form of the anode current on full load by a rectangle (dotted lines in fig. 7) whose height is equal to the maximum value of $I_{a \max}$ of the anode current and the width such that the area of the rectangle is the same as that of the actual wave-form of the current. This width is taken to be k times the duration of a cycle, k being usually 0.20 to 0.25.

As a further approximation we assume that, when current is flowing, the average anode voltage is equal to the voltage V_{ak} at which the bend in the curve occurs. The oscillatory circuit thus receives an amount of energy $kI_{a \max}(V_B - V_{ak})$ per cycle or, expressed in power, $f k I_{a \max}(V_B - V_{ak})$. The power delivered by the oscillator is $I_{0 \max} V_{0 \text{ nom}}$. The efficiency will be denoted by η_a , disregarding the heater and screen-grid-current; hence:

$$\eta_a f k I_{a \max} (V_B - V_{ak}) = I_{0 \max} V_{0 \text{ nom}}$$

This equation may be employed to estimate $I_{a \max}$ for the purposes of the design when $I_{0 \max}$, $V_{0 \text{ nom}}$ and V_{ak} are known, values for V_B and f are decided upon and η_a and k can be evaluated in accordance with experience. Practical values of k have already been indicated; η_a usually lies between 0.3 and 0.6, dependent on the value of $\tan \varphi$ and the quality factor of the coil (see above).

The value of $I_{a \max}$ obtained in this manner may exceed the rated value for the valve in question, but in this case two or more valves can be employed, either in parallel or, if an even number, in push-pull.

Reducing the internal resistance

Although the above-mentioned methods of limiting the amplitude of the alternating voltage already considerably assist in keeping down the internal resistance of the generator, they are not generally adequate if requirements as laid down are very stringent.

In such cases automatic voltage control is applied. The H.T. voltage is then regulated by a control voltage which is varied automatically in such a way that the high tension remains reasonably constant on varying loads.

Voltage control of this kind in no way minimizes the usefulness of the methods of avoiding variations in the output already referred to (choice of φ_{\max} , limitation circuit in the oscillator), for, if such precautions are neglected, a possible defect in the automatic control might cause the voltage to exceed the safe limits.

There is a choice of methods of obtaining the necessary control voltage, viz. this voltage may be:

- proportional to the high tension or to the difference between this and a constant reference voltage;
- proportional to the output current of the generator;
- made up of the sum of two voltages, one of which is obtained as in (a), the other as in (b).

Method (a) has the advantage that it ensures high stability, but it does not admit of making the internal resistance zero or negative. This is possible with method (b), but here there is some risk of instability. The combined method (c) frequently affords a satisfactory combination of the advantages of (a) and (b).

Similarly, the control voltage can be made to operate on the high tension in different ways, and examples will now be given of two such methods as applied to 50 kV generators made at the Philips Laboratories at Eindhoven.

- The first of the control circuits described in the following refers to a 50 kV generator giving a

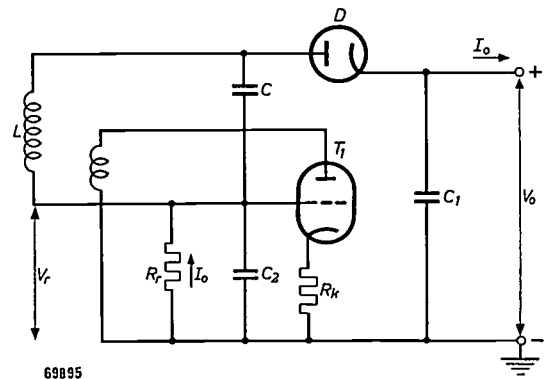


Fig. 8. Circuit of H.T. generator in which the voltage V_0 is maintained at a constant level by a "damper" valve T_1 . R_r = resistor across which the control voltage $V_c = I_0 R_r$ occurs. C_2 = smoothing capacitor. R_k = bias resistor. For meanings of the remaining symbols see figs. 2 and 4.

maximum mean current of 0.45 mA. The control voltage is obtained by passing the current I_0 through a resistor (R_r , fig. 8), that is, in accordance with method (b) above. This voltage is part of the grid potential of a triode, the anode circuit of which is coupled to the oscillatory circuit $L-C$ of the H.T. generator. The triode, or "damper" valve, exercises control on the voltage by reason of the amount of damping which it affords, dependent on whether the grid voltage is high or low. When the load is reduced, and with it the corresponding part of the total damping of the oscillator, $|V_r|$ is so reduced that the valve in question provides an equivalent amount of damping, the overall damping thus remaining constant.

An advantage of this arrangement is that the power taken by the oscillator from the source of supply is constant. Should the internal resistance of this source be on the high side, it would have no effect on the internal resistance of the H.T. generator. It is of course a drawback that the extra valve must be capable of dissipating considerable power, namely $V_{0nom} \times I_{0max}$ (in our example 22.5 W), for which reason this method is suitable only in cases where the power generated is not very high.

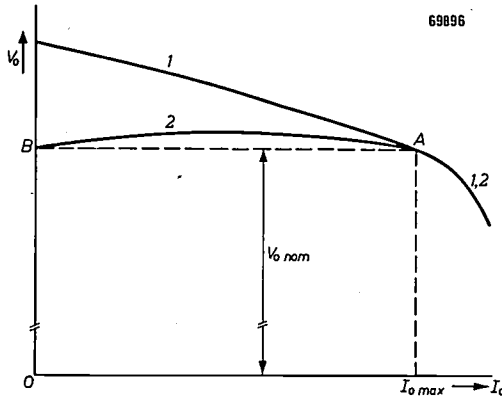


Fig. 9. Curves representing $V_0 = f(I_0)$ of an H.T. generator: 1 without, 2 with automatic voltage control.

The adjustment of the circuit is quite simple. In fig. 9 (1) shows the curve $V_0 = f(I_0)$ without voltage control; the curve is such that the nominal voltage V_{0nom} is available (point A) when the maximum current I_{0max} flows. Let us now see what happens when the control becomes operative. Under no-load conditions no current passes through the resistor R_r (fig. 8), in which case the voltage between grid and cathode of the damper valve is only as provided by the biasing resistor R_k . This resistor ensures a bias that will make the no-load voltage of the generator equal to V_{0nom} ; the triode used for the damping then introduces just as much damping in the oscillatory circuit as the rectifier does on full load. The curve showing $V_0 = f(I_0)$ with control operating (2), thus passes through the point B (fig. 9), which is level with A.

When the generator is placed under load, current flows through R_r , the bias is increased and the damping effect of the triode is reduced. The value of R_r should be so adjusted, that the triode just fails to pass current and therefore affords no damping when the generator is fully loaded. Curve 2, then, also passes through the point A.

In the ideal case, the characteristic would be the straight line BA. However, since not all effects in the circuit are linear functions of the control voltage V_r , the characteristic is usually somewhat curved, as in 2, fig. 9. On small loads the internal resistance of the generator is then negative (increasing voltage with increasing current). So long as the absolute value of this resistance is not too high, there will be no instability, owing to the positive resistance of the load.

Let us for a moment suppose that the triode is replaced by a resistance equal to the average internal resistance R_i of the triode, at the grid bias that occurs when the generator is not under load ($V_r = 0$). Taking into account the fact that

the triode passes current every other half cycle only, the power dissipated will be:

$$W_a = \frac{1}{2} \frac{V_a^2_{rms}}{R_i}$$

where $V_{a rms}$ is the r.m.s. value of the alternating voltage in the anode circuit. Since W_a is known ($= V_{0nom} \times I_{0max}$), we thus also know the appropriate value of $V_{a rms}$ to meet our requirements.

In order to avoid dissipating the whole of the power by the triode, a resistor may be placed in series with this valve to absorb some of the power. Nevertheless, it may be found necessary to use two or more valves in parallel or push-pull.

2) The second method of providing automatic control, which we shall now discuss⁹⁾, was incorporated in a generator for 50 kV, 1.5 mA. Owing to the amount of power involved (75 W), the use of a damper valve is no longer suitable.

The control voltage obtained by the above-mentioned combined method (c) operates on the control grid of the oscillator valve. The circuit is reproduced in fig. 10.

The current I_0 is passed through a resistor R_r ; the higher the current, the greater the grid bias applied to a small triode T_2 . The potential of the anode (point x) reproduces in an amplified form the fluctuations in the voltage $I_0 R_r$ with the correct

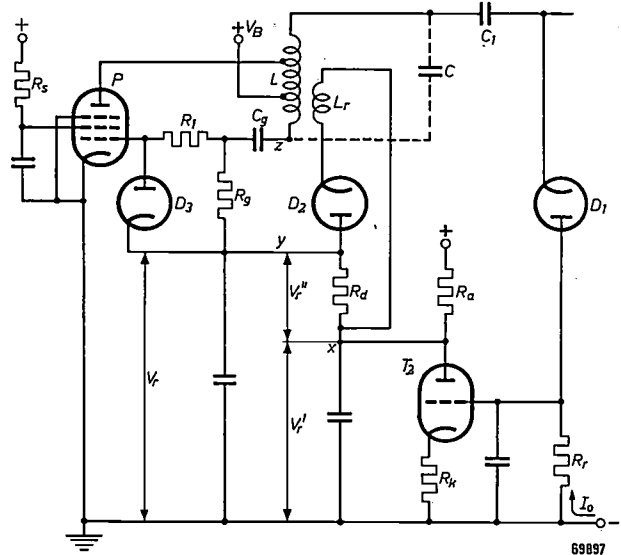


Fig. 10. Circuit of H.T. generator with stabilised voltage, with control voltage V_r applied to the grid of the oscillator valve P.

V_r consists in part of V_r' , which depends on the current I_0 , and in part of V_r'' , which is proportional to the voltage V_0 . $T_2 =$ auxiliary triode. $D_2, D_3 =$ auxiliary diodes. $R_a =$ load resistor. $R_d =$ resistor across which the voltage V_r'' is produced by the rectified current from D_2 . $R_1 =$ resistor for flattening the positive peaks of the alternating voltage at the grid of P. Of the cascade rectifier only the first capacitor (C_1) and first diode (D_1) are shown. For the remaining symbols see figs. 2 and 4.

⁹⁾ Developed by L. J. van de Polder of the Research Laboratory, Eindhoven.

sign for the voltage V_r' between x and earth to be used as control voltage for the grid of the oscillator valve. The level of the potential at x is too high, however, and a second potential, of opposite polarity, is used in conjunction with the first. This latter potential is obtained by rectifying (diode D_2) the alternating voltage induced in a winding L_r which is coupled to the coil L . The direct voltage component V_r'' then occurring across R_d is of a value and polarity that will give the point y (fig. 10) a negative potential with respect to the cathode of P (earth).

The action of the circuit is as follows. On increasing loads, $I_0 R_r$ rises, the anode current of T_2 drops, the potential at point x increases and, assuming V_r'' to be constant, that of y rises to an equal extent. As a result, the oscillatory circuit receives more power to meet the increased load.

Actually, V_r'' is not constant, however, but is proportional to the alternating voltage across L_r , that is, to the alternating voltage across the L - C circuit, and, therefore, also to the voltage V_0 to be controlled. This proportionality is to our advantage, since negative feedback occurs by way of V_r'' , which counteracts variations in V_0 and thus promotes stability.

The circuit is so adjusted that on full load the point y is just at earth potential (it must not be at a higher potential as this precludes any controlling action), and that on full load the potential of y will have dropped to the extent that the L - C circuit receives just enough power to produce the same high tension V_{0nom} as on full load. This is easily arranged by employing suitable values of R_r , R_k and R_a (fig. 10).

To mention one or two details of the circuit shown in fig. 10, the diode D_3 , in conjunction with resistors R_g and R_1 ensures that the control grid assumes the potential of point z when the positive peaks of the alternating voltage occur at the point z . This potential will then determine the amount of power applied to the L - C circuit.

Resistor R_1 ensures that the alternating voltage component at the control grid shall not be sinusoidal (as at x), but that it takes the form of flattened peaks, thus improving the waveform of the anode current of the pentode.

Two practical examples

In conclusion, particulars will now be given of the two 50 kV generators constructed in the laboratory for the purpose of carrying out tests with large-screen projection television (approx. 3 m \times 4 m).

50 kV, 0.45 mA

This generator, which has air-insulation and which employs the principle of constant damping, is depicted in fig. 11. The rectifier (on the right in the photograph) consists of four stages in cascade, each using a DY 30 diode. Each of the valve holders in the rectifier is enclosed in a well rounded-off

brass ring to avoid corona. The valves and capacitors of the rectifier are enclosed between plates of polystyrene.

The filaments of the diodes, which consume only 0.25 W each, are fed with high-frequency current supplied by the oscillator; the filaments of the first and second diode are each fed from a specially insulated winding coupled direct to the main coil, but the feed of the third and fourth diodes is taken from a separate insulating transformer in order to

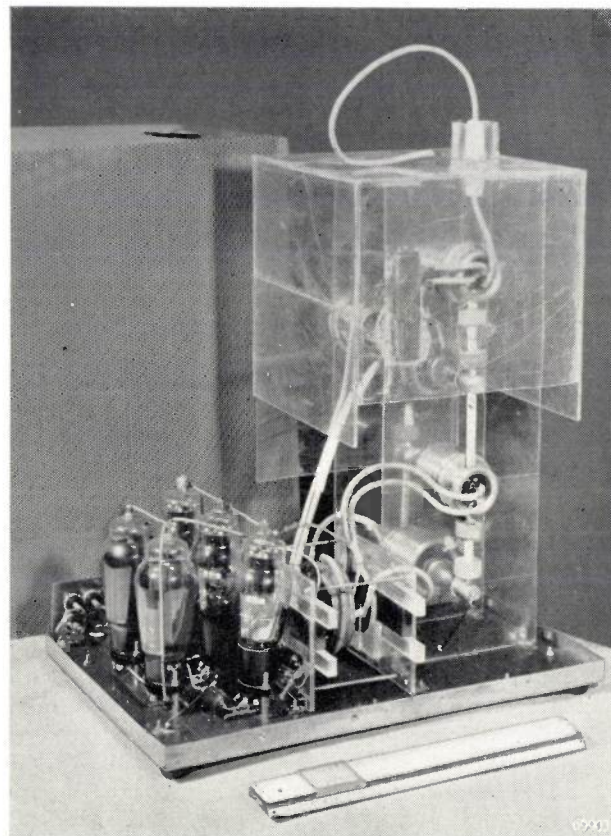


Fig. 11. Experimental 50 kV, 0.45 mA H.T. generator. From left to right the two damping valves in parallel, the three oscillator valves in parallel, the oscillator coil and, between polystyrene plates, the rectifier (four diodes DY 30 in cascade).

divide the wide difference in potential between these filaments and earth. The transformer consists of a closed magnetic circuit of Ferroxcube carrying a primary and two secondary windings, each of which comprises only a few turns of insulated wire. The oscillator frequency is 24 kc/s.

Fig. 12 shows the measured characteristic $V_0 = f(I_0)$; with no load, V_0 is only 2% above the nominal value of 50 kV. The figure also shows the total direct anode current I_a taken from a 500 V source; because of the automatic control on the damping, this current — of roughly 120 mA — is practically independent of I_0 .

The efficiency on full load is found to be:

$$\eta_a = \frac{50\,000 \times 0.45 \times 10^{-3}}{500 \times 120 \times 10^{-3}} \times 100\% = 37.5\%$$

50 kV, 1.5 mA

Fig. 13 shows the 50 kV, 1.5 mA generator,

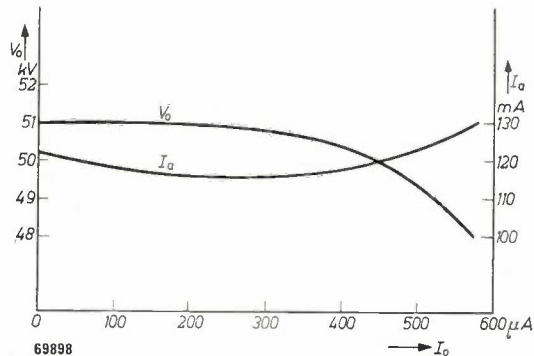


Fig. 12. D.C. voltage V_0 , input and total anode current I_a drawn from a 500 V source, as functions of the load current I_0 , of the generator shown in fig. 11.

also in experimental form. In this case the oscillator coil and cascade circuit are immersed in transformer oil, thus making it possible to employ smaller components (paper capacitors; EY 51 diodes); the complete unit accordingly gains in compactness.

The cascade circuit comprises six stages mounted in a circle around the coil, the voltage increasing in a clockwise direction. This arrangement was adopted in order to ensure that the leads to the diodes, the heaters of which are again fed from the oscillator, would be as short as possible. The oscillator coil is of a conventional type with oiled-paper insulation and Ferroxcube core; the frequency is about 30 kc/s.

On the chassis in the upper left-hand corner of the photograph will be seen the four oscillator valves PL 81, connected in push-pull and parallel, together with the triode, diodes and other components of the control circuit, the basic diagram of which is reproduced in fig. 10.

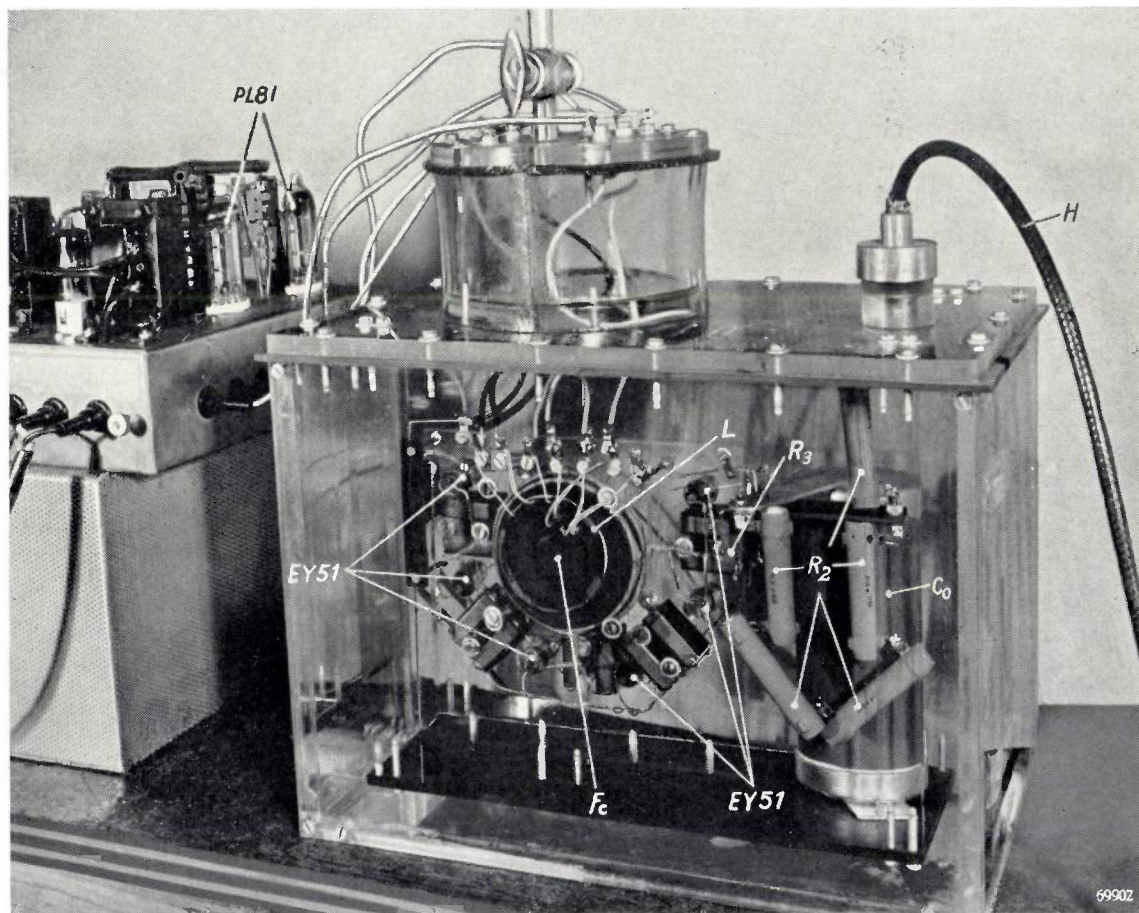


Fig. 13. Experimental H.T. generator to deliver 50 kV, 1.5 mA. Left-hand top: chassis with oscillator valves (PL 81) and voltage control circuit. Centre: oil-filled container with oscillator coil L (and Ferroxcube core F_c) and the six cascade stages of the rectifier (with diodes EY 51), mounted round the coil; reservoir capacitor C_0 , protecting resistors R_2 and R_3 . H is the H.T. cable for connection to the cathode-ray tube. The container is made of "Perspex", so that the internal components can be observed.

Fig. 14 shows the characteristics of the output voltage V_0 and the total anode current I_a derived from a 1000 volt direct-voltage source, as functions of the load current I_0 . In contrast with the first example, the current I_a rises on an increasing load.

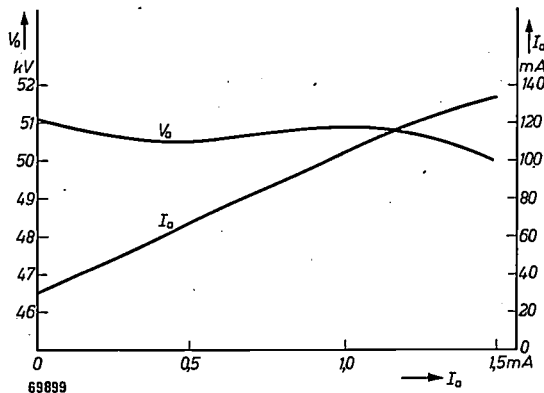


Fig. 14. Direct voltage V_0 , and total input anode current I_a drawn from a 1000 V source, as functions of the load current, I_0 , of the generator shown in fig. 13.

On full load $I_a = 135$ mA, and the efficiency is then:

$$\frac{50000 \times 1.5 \times 10^{-3}}{1000 \times 135 \times 10^{-3}} \times 100\% \approx 56\%.$$

The voltage control circuit already described ensures that the internal resistance (measured statically) will be sufficiently low. In television pictures, however, fairly large bright areas may occur in which the beam current would greatly exceed the value of $I_{0 \max}$ and, in order to prevent the acceleration voltage from dropping to too low a level, it is advisable to connect a reservoir capacitor to the output terminals. Seeing that exceptionally high standards were set for the picture quality of the projector for which this H.T. generator was designed, the actual reservoir capacitor

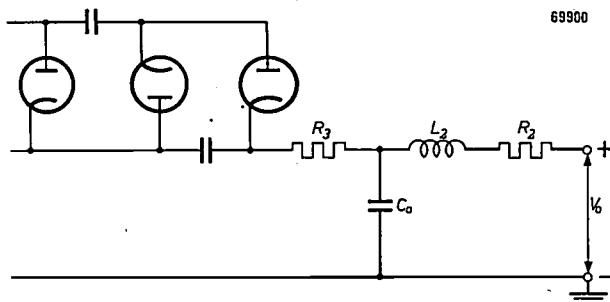


Fig. 15. Circuit of the reservoir capacitor C_0 and safety resistors R_2 and R_3 in the generator depicted in fig. 13. $L_2 =$ self-inductance of the resistor R_2 . On the left: part of the cascade circuit.

used had a fairly large value (10,000 pF), but in many cases a smaller capacitance would suffice.

The large quantity of energy stored in this capacitor would prove detrimental in the event of breakdown in the cathode-ray tube which, albeit sporadically, is liable to happen. To prevent possible damage, a resistor R_2 is provided (figs. 13 and 15). This is a wire-wound resistor the self-inductance of which (0.5 mH) is sufficient to limit the current from the reservoir capacitor to some tenths of an ampere in the event of a flashover of very short duration (10^{-8} sec). If the duration is longer its resistance of 50,000 Ω limits the current to 1 A. A second resistor ($R_3 = 0.5$ M Ω) serves to limit the charging current of the capacitor C_0 when the circuit is switched on, thus protecting the diodes. The combination of R_3 and C_0 moreover acts as a smoothing filter.

The effect of the resistor R_2 can best be demonstrated by connecting a spark-gap to the generator; without the resistor R_2 each flashover produces a sharp crack, whereas with the resistor only a faint click is heard. The experimental generator proved well able to withstand such tests.

In view of the efforts made to keep the internal resistance as low as possible it may appear strange to increase the resistance by as much as 0.5 M Ω , but this value is virtually low in comparison with the internal resistance that would be encountered in the absence of any voltage control (about 10 M Ω). The extra resistance R_3 therefore has little effect on the final result.

The development of this immersed generator has been a source of valuable experience against the time when units for even higher voltages may be required.

Summary. Television pictures of large dimensions having at the same time high brightness and definition can be obtained only by means of tubes operating with very high acceleration voltages. Present-day receivers working on the projection principle employ 25 kV, but equipment developed in the Philips Laboratories at Eindhoven producing a picture of 3 m \times 4 m uses 50 kV. Four methods of generating such high voltages are reviewed and one of these is described in detail, namely one in which an oscillator, preferably having a coil with Ferroxcube core, delivers an alternating voltage of 20 to 30 kc/s, this being rectified by a cascade circuit. This proves to be the best arrangement to meet the conditions imposed on an H.T. generator for the purpose in question, namely constant direct voltage from no-load to full-load conditions, with a rapid fall in potential when full load is exceeded. The rectifier and oscillator in such systems are discussed, as also methods of ensuring a low internal resistance. Examples are given in relation to two experimental generators (50 kV, 0.45 mA and 50 kV, 1.5 mA).

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

ELECTRONIC INTENSIFICATION OF FLUOROSCOPIC IMAGES

by M. C. TEVES and T. TOL.

621.386.8 : 616-073.75 : 621.383.8

In essence, the generation of X-rays, which entails the production of a beam of electrons in a vacuum tube, falls within the technique of electronics. It is a noteworthy point that in X-ray diagnostics the subsequent stage, that is, the conversion to a visible picture of the X-rays which have passed through the object to be examined, is now also tending to become dependent on electronics for a satisfactory solution. In this context it is found that the number of X-ray quanta available virtually sets a limit to the amount of detail in the information that can be derived from fluoroscopy.

The problem of screen brightness in fluoroscopy

X-rays have the capacity to damage or destroy human tissue. In X-ray therapy, use is made of this fact by exposing morbid tissue to these rays for such periods and at such intensities that in total it will receive a certain, heavy, dose. In X-ray diagnostics, on the other hand, care must be taken that patients receive the smallest possible dose during examination so that they will not suffer any injurious consequences, even as a result of repeated examinations.

Using the conventional method of fluoroscopy, therefore, the radiologist must do the best he can with an X-ray image of very low brightness. In fact, for a given duration of an examination (a few minutes at the most), limitation of the total dosage means limitation of the dosage rate to which the patient may be subjected and, accordingly — dependent upon the absorption of the rays by the patient — limitation of the brightness that can be obtained from the screen. In gastric fluoroscopy, for example, with the X-ray tube operating on a voltage of, say, 90 kV and a current of 2 mA, the image obtained on a high-quality screen has an average brightness of about 0.003 cd/m². For comparative purposes it may be said that the average brightness of a landscape by the light of the full moon is 0.01 cd/m².

In order to be able to make a good visual assessment at such low brightness values, the radiologist

must first remain in darkness for at least 15 minutes, so that his eyes may be adapted to the low brightness level. Even so, his visual acuity and contrast sensitivity¹⁾ are then fairly low. Hence, for the diagnosis of conditions involving slight abnormalities which often reveal only a low contrast, this is a very inconvenient limitation. It will therefore be easily appreciated that, since the inception of X-ray diagnostics, higher image brightnesses have been an ever-present need.

If more information is to be obtained from the fluoroscopic screen and at the same time a solution found for the problem of the necessity for dark adaptation of the eyes, an increase in image brightness to the extent of at least a factor of 100 to 1000 is needed, as pointed out by Chamberlain in 1942²⁾. It would not be possible to achieve this simply by increasing the efficacy of existing types of screen; these convert roughly 5% of the X-rays falling on them to visible radiation, and this radiation has approximately the most advantageous wavelength (i.e. in the yellow-green part of the spectrum, where, for a given strength of the radia-

¹⁾ Visual acuity is defined as $1/a$, where a is the angle subtended by the finest distinguishable detail. Contrast sensitivity is defined as $1/c$, where c is the lowest distinguishable contrast $\Delta B/B$; in this, B is the brightness of the background and $B \pm \Delta B$ the brightness of an object seen against that background.

²⁾ W. E. Chamberlain, *Fluoroscopes and fluoroscopy*, *Radiology* 38, 383-413, 1942.

tion, the strongest impression of brightness is received by the eyes). Thus, the gain that is theoretically possible in this way would not exceed a factor of 20.

In recent years several methods, using electronic aids, have in fact been devised for intensifying the brightness of the fluoroscopic image to a considerable degree³⁾. In the present article we shall describe an image intensifier designed in the Philips Laboratories at Eindhoven. This equipment is based on the well-known principle of the image converter, another development of this laboratory, a description of which was given as early as in 1934⁴⁾. The fluoroscopic images obtained with the aid of this instrument are roughly 1000 times brighter than those produced on the ordinary fluoroscopic screen.

Description of the intensifier

The more important details of the unit are illustrated in *fig. 1*. A fluoroscopic screen on a support of thin aluminium is mounted in an evacuated glass tube. A photocathode lies in contact with the screen. When X-rays pass through the glass wall of the

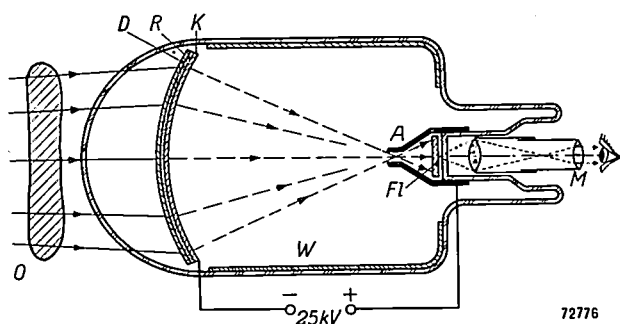


Fig. 1. Diagrammatic cross-section of the image intensifier. *R* = fluoroscopic screen on which the X-rays are received after passing through the object to be examined *O*, and the glass wall of the tube. *D* = support of fluoroscopic screen and photocathode *K*. The fluorescent radiations produced in *R* release electrons from the photocathode *K*. The "electronic image" is reproduced on a reduced scale by the electric field between *K* and the perforated anode *A*, on a viewing screen *Fl*; this image is observed through a simple microscope *M*. *W* is a conductive layer on the inside of the tube.

- ³⁾ J. W. Coltman, Fluoroscopic image brightening by electronic means, *Radiology* 51, 359-367, 1948.
 R. J. Moon, Amplifying and intensifying fluoroscopic images by means of a scanning X-ray tube, *Science* 112, 389-395, 1950.
 R. H. Morgan and R. E. Sturm, The John Hopkins fluoroscopic screen intensifier, *Radiology* 57, 556-560, 1951.
 P. C. Hodges and L. Staggs, Electronic amplification of the röntgenoscopic image, *Amer. J. Röntgenology Rad. Ther.*, 66, 705-710, 1951.
- ⁴⁾ G. Holst, J. H. de Boer, M. C. Teves and C. F. Veenemans, An apparatus for the transformation of light of long wavelength into light of short wavelength, *Physica* 1, 297-305, 1934.

the tube to the screen, this fluoresces. The fluorescent light sets electrons free from the photocathode; at any point on the photocathode the number of photo-electrons released per second is proportional to the luminous intensity of the fluoroscopic screen at that point. The luminous image with all its variations in brightness is thus transformed into an "electronic image" with corresponding variations in current density. By means of an electric field of a certain configuration (electrostatic "electronic lens") this electronic image is reproduced on a second fluorescent screen, the viewing screen, reduced 9 times in size. The energy of the electrons falling on this screen is in part converted to fluorescent light, and a reproduction of the image on the fluoroscopic screen is thus seen on the viewing screen reduced 9 times. This reduced image is viewed through a simple microscope of about $9 \times$ magnification, and finally the fluoroscopic image is thus seen erect and in its original size (both the electronic lens and the microscope produce inverted images).

The image seen through the microscope is thus identical with that seen on an ordinary fluoroscopic screen; it is, however, about 1000 times brighter, assuming that the image intensifier and the fluoroscopic screen both receive X-rays of the same intensity.

How is the brightness intensification obtained?

Two factors (which are not in fact independent of each other) contribute towards the intensification of the image brightness.

The first and more readily understood factor is an increase in the total luminous flux. The electrons emitted by the photocathode are not only focused on the viewing screen by the electrical field, they are also accelerated by it; in the particular instrument under review a potential of 25 kV is applied between the viewing screen and the photocathode. The more energy possessed by an electron on arrival at the screen, the more fluorescent light it will produce. Notwithstanding the fact that only about 1 in 10 of the light quanta from the fluoroscopic screen liberates a photo-electron and that only some 10% of the electronic energy is converted to light by the viewing screen, the last mentioned screen — owing to the energy imparted to the electrons — is able to yield a total of 10 to 15 times as much luminous flux (in lumens) as an ordinary fluoroscopic screen, for the same object viewed.

The second, more important factor is the gain produced by the above-mentioned electron-optical

reduction of the size of the image. All the electrons coming from the photocathode contribute towards the final image, and the energy that each electron is capable of converting to light after passing through the electron-optical system is independent of whether the electrons are distributed over a large or over a small area. By employing a reduction of 9 times we concentrate the electrons in an area 9^2 times smaller, and the total luminous flux is thus emitted from an area that is about 80 times smaller than would be the case in reproduction on a scale of 1 : 1. This, by definition, represents an increase in brightness⁵⁾ by a factor of 80.

The total gain in brightness is equal to the product of the "lumen gain" and the gain due to the reduction in size, i.e. in the tube in question, 10 to 15 times $9^2 = 800$ to 1200.

It is perhaps not superfluous to add that the gain brought about by the electron-optical reduction cannot be obtained by the ordinary methods of optical reduction. According to Abbe's law, the apparent brightness of an object does not vary as a result of the introduction of any kind of optical instrument between the object and the eye of the observer, that is, if such an instrument can be regarded as ideal, without the absorption and reflection losses that normally occur in lenses and mirrors (and assuming the exit pupil of the instrument to be at least as large as the pupil of the eye).

This will readily be appreciated on reference to *fig. 2a*. Let the brightness of the object be B , and the solid angle subtended by the lens when seen from the object ω . The lens then receives from a surface element df a luminous flux of $\Phi = B\omega df$. This luminous flux is concentrated in a smaller area df' in the image space. This area thus again emits the luminous flux Φ , this time, however, within the greater solid angle ω' (or, with diffusion, in an even larger angle). The brightness of df' is $B' = \Phi/\omega'df'$ and, since we know from the laws of optics that $\omega'df' = \omega df$ (the sine law), B' is equal to B .

The difference that is seen to exist in this case by no means implies that the analogy between electron-optics and light-optics is no longer valid. Abbe's law as stated in the above simplified form holds good only when the object and the image are located in media having the same refractive index, and in ordinary optical instruments this is usually the case. If the refractive index of the medium containing the image is higher than that of the object, the brightness of the image can certainly be greater than that of the object itself. This is exactly what happens in electron-optics. The potential field in which the electrons are moving may be regarded as a medium the refractive index of which is continuously variable; in fact, the index is at any point propor-

tional to the velocity of the electrons. The refractive index at the site of the viewing screen is therefore about 100 times higher than that at the photocathode.

The analogy is thus not only restored formally, but may again be illustrated in a simple way. Owing to the continuous variability of the refractive index, the "rays" are now not straight, but curved. The path of the rays in the electrostatic electron-lens therefore takes the form shown in *fig. 2b*, in contrast with *fig. 2a*; it will be seen that the flux radiated from the surface element df within a solid angle almost equal to 2π is employed and is concentrated in a much smaller solid angle at the surface df' .

By reason of the influence of the velocity of the electrons, it is seen that the "lumen intensification" and the brightness gain due to the reduction in size virtually spring from the same cause, i.e. the presence of the accelerating field.

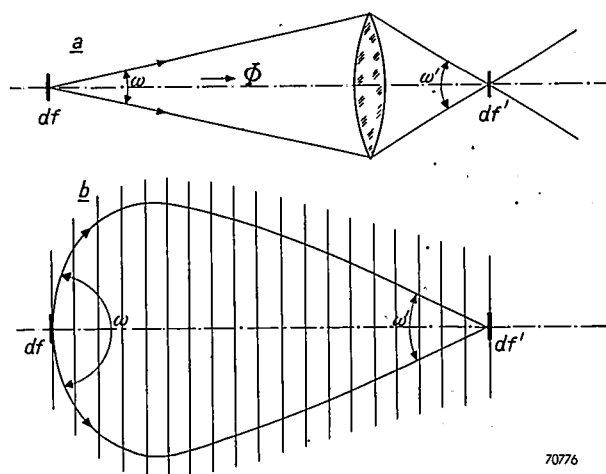


Fig. 2. a) Path of rays on optical reproduction of a surface df , at df' (illustrating Abbe's law).

b) The same, as applicable to electron-optical reproduction with an electrostatic lens, as in the image intensifier.

In the light of Abbe's law it will now be understood why in the last stage of the image intensifier, viz. the optical enlargement by the microscope of the image on the viewing screen, the gain in brightness arising from the electron-optical reduction is not neutralised (provided that the microscope meets the condition as regards the exit pupil to which reference has already been made).

It may possibly strike the reader that in the foregoing, particularly in our explanation of the lumen gain, the luminous flux of the viewing screen is each time compared with that of an ordinary fluoroscopic screen and not, as might appear more obvious, with the luminous flux of the fluoroscopic screen in the image intensifier itself. The reason is that the last-mentioned screen need not be observed and that the luminosity curve of the fluorescent light emitted by it may therefore be such as to be quite unsuitable from the point of view of direct observation; in fact the fluorescent

⁵⁾ According to a recent standardisation decision on lighting technology terms, we should speak here of luminance.

radiation might even be entirely invisible, for instance ultra-violet or infra-red, provided only that the photocathode be sensitive to such radiation. Our image intensifier is equipped with a photocathode the quantum efficiency of which is highest (about 10%) when it is exposed to blue light. A phosphor giving blue fluorescence was accordingly chosen for the fluoroscopic screen. For the viewing screen, on the contrary, if a choice from different materials of the same luminous efficiency is available, that colour should be employed of which the fluorescent spectrum coincides as nearly as possible with the maximum in the spectral sensitivity curve of the eye. This would be a phosphor emitting yellow-green light. Clearly, if we compared the brightness of such a viewing screen with that of the blue emission of the fluoroscopic screen in the image intensifier, we would calculate a much greater brightness intensification value than 1000 times, but at the same time one that would have no practical significance.

For this reason it would be as well to dwell for a moment on the precise meaning of the name "image intensifier" or "brightness intensifier". In effect the instrument is an image converter, which converts the "image" of the invisible X-rays to a visible image, by analogy with the well-known instruments for transforming infra-red radiations into visible images (see the article referred to in note 4)) which are employed in numerous different forms. The fact that this conversion in our case involves as intermediate stage a visible fluorescent image (as in normal fluoroscopy) is attributable only to the impossibility of liberating photo-electrons direct from a photocathode with any degree of efficiency (and with not too high an initial velocity) by means of the X-rays themselves. Reasonable efficiency in this release of electrons is naturally of the highest importance in the fluoroscopic image intensifier, as otherwise the effect of the subsequent gain would be lost and we might then just as well employ only the "image conversion" as obtained from the ordinary fluoroscopic screen.

Details of construction

The blue fluorescence of the fluoroscopic screen as matched to the photocathode is obtained from zinc sulphide activated with silver, and the yellow-green fluorescing viewing screen is prepared from very finely divided zinc sulphide-zinc selenide. The effective diameter of the cathode face is 13.5 cm, this thus representing the maximum size of the object to be examined. The image on the viewing screen is only 1/9th of this size, viz. 1.5 cm across.

The electrical field producing the electron-optical image is generated basically by means of only two electrodes, one of which (the cathode) is formed by the concave spherical surface of the combined fluoroscopic screen and photocathode, as

well as by a conductive layer on the inside wall of the glass tube. The other electrode (the anode) is also spherically curved and is concentric with the photocathode; it has in it an aperture to allow the electrons to pass to the viewing screen behind it⁶⁾. Between the cathode and the anode a potential of 25 kV is applied.

This simple design, employing only two electrodes, has proved to be very effective. Definition in the image is ample and the field strength on the only slightly curved photocathode is very low, this being an asset for a reason which is given in another paragraph. When the tube was still in the experimental stage, the configuration offered sufficient latitude, viz. in the radii of curvature of cathode and anode, allowing for example a variation in the amount of reduction of the image. For practical purposes it was later found to be better for the inner lining of the tube to function as a separate electrode and to maintain it at a slightly positive potential with respect to the photocathode. Without this, only the electrons emitted from the central part of the photocathode pass through the opening in the anode, the peripheral parts thus not being reproduced. With the auxiliary potential applied the peripheral electrons also pass through the anode, the effective image area of the fluoroscopic screen being thus almost doubled. At the same time, the slight "pin-cushion" distortion of the image that occurs without the auxiliary voltage is in this way almost entirely eliminated. Another important advantage of the auxiliary potential — which is variable — is that by means of it the image on the viewing screen can be adjusted for maximum definition whilst the tube is in use. Consequently, the tolerances in the electrode spacing, during the manufacture of the glass tube, can be relatively wide.

The 25 kV applied between cathode and anode may be obtained for instance from the well-known small high-tension generator developed in this Laboratory for projection television⁷⁾. This generator also supplies the necessary voltage for the inner tube coating. The cathode is earthed, so that only one of the three poles of the image intensifier need be suitable for carrying high voltage. Only a very small amount of power is required to operate the tube, the photo-current for normal fluoroscopy being in the order of 10^{-9} A. For this reason and also in view of the fact that in this case no stringent con-

⁶⁾ For a more detailed description of the principle of this two-electrode lens the reader is referred to an article by P. Schagen, H. Bruining and J. C. Francken, Philips Research Reports 7, 119-130, 1952 (No. 2).

⁷⁾ G. J. Siezen and F. Kerkhof, Philips Techn. Rev. 10, 125-134, 1948.

ditions need be imposed as regards voltage stability (a basic advantage in purely electrostatic electron-optical systems), an even simpler high-tension unit would suffice.

Fig. 3 depicts the complete image amplifier tube. It is cylindrical, the diameter being 17.5 cm; the over-all length, including the small microscope, is 45 cm. To protect the tube against jarring and to afford local protection from stray X-rays, the tube proper is contained in a lead-lined aluminium jacket which can be mounted quite simply on a stand; see fig. 4. Tube and jacket together weigh 7.5 kg approx.

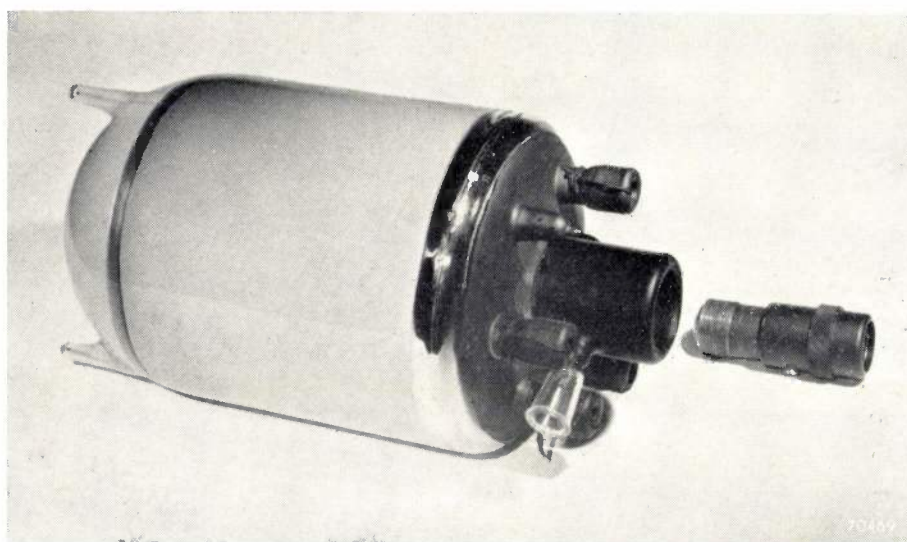


Fig. 3. The Philips X-ray image intensifier, producing an image that is approximately 1000 times brighter than that of the ordinary fluoroscopic screen. The tube is 45 cm in length, 17.5 cm in diameter. The effective area of the screen is a circle 13.5 cm in diameter. On the right will be seen the viewing end into which the microscope shown alongside is fitted.

Definition and contrast in the image

When the image intensifier is in use it is not necessary for the radiologist to allow time for his eyes to adapt themselves; examinations can be made in a room with almost normal illumination. This is in itself a great practical advantage that will almost certainly outweigh the drawback that the image cannot be surveyed with the naked eye, but only through an optical system. Moreover, it may be expected, as explained in the introduction, that the radiologist will be in a position to perceive smaller details than is possible with the ordinary fluoroscopic screen because of the higher visual acuity and contrast sensitivity of the eye at the increased brightness level. In order that this expectation — which we shall enlarge upon presently — may be realised, it is an essential condition that the blurring and loss of contrast introduced by the image

intensifier itself, as by any other form of optical instrument, shall be kept within certain limits.

Definition

In normal fluoroscopy the definition of the image will be determined by the finite dimensions of the focus of the X-ray tube (half-shadow width or geometrical blurring) as well as by the thickness and grain size of the fluoroscopic screen (intrinsic or screen blurring). Screen blurring is usually about 0.7 mm, sometimes less, down to 0.4 mm; the geometrical blurring is generally slightly less. The resulting definition in the fluoroscopic image

is thus somewhat less than 1 mm (the total blurring being equal, approximately, to the square root of the sum of the squares of the geometric blurring and the screen blurring⁸⁾).

Screen blurring can be reduced considerably by using thinner screens (and finer grain). This is not done in the case of ordinary fluoroscopy, as such screens would transmit a larger amount of X-rays, which are thus ineffective, resulting in reduced screen efficiency⁹⁾. Moreover, better definition would be

⁸⁾ See H. A. Klasens, The blurring of X-ray images, Philips Techn. Rev. 9, 364-369, 1948.

⁹⁾ As far as efficiency is concerned there is an optimum thickness, for in screens that are too thick, apparently too much absorption of the emitted light will take place in the screen itself. (This conflict between useful absorption of the incident radiation and wasteful absorption of the rays emitted in the image intensifier again occurs in the photocathode and in the viewing screen.)

of little use, since, at the low brightness levels of normal fluoroscopic screens, visual acuity is in any case too low for the perception of finer detail.

In the image intensifier the position is very different. Here, too, there is a primary lack of definition in the image owing to the intrinsic blurring in the fluoroscopic screen employed. However, there is now no objection to the use of a relatively thin screen, thus sacrificing some of the gain in brightness to benefit the definition.

The grain size of the viewing screen used is very fine and the screen itself very thin (the penetration depth of electrons at 25 kV is about 12μ), so that the intrinsic blurring in this screen, referred to natural image size, is less than that of the fluoroscopic screen. Blurring by the electron-optical reproduction system is at least one order of size less (see formula (26) in the article referred to in note ⁶) and can therefore be ignored, as can, also, the limitation in the resolving power of the microscope

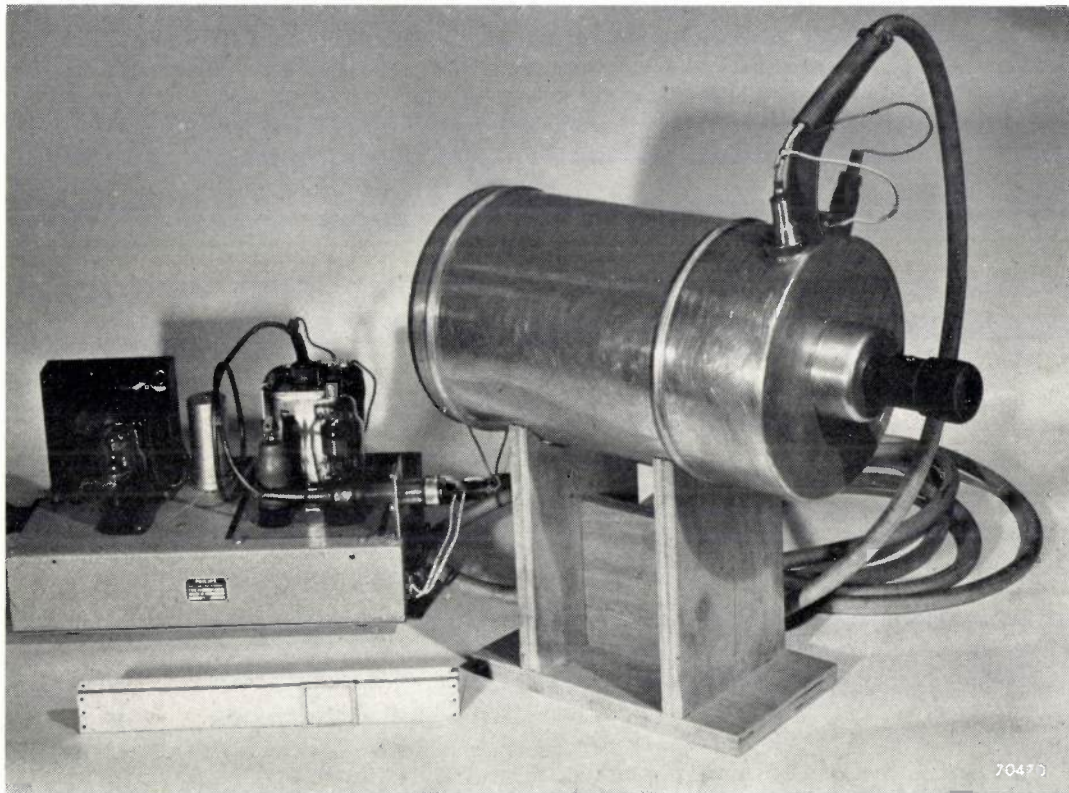


Fig. 4. The complete equipment for fluoroscopy by means of the image intensifier. The tube is housed in a lead-lined aluminium jacket and the whole can be conveniently mounted on a stand. Left: the small high-tension unit for supplying the necessary 25 kV, D.C.

A real advantage is obtained by so doing, for the visual acuity at the brightness level of the image intensifier is so high that the eye can reap the benefit of the increased definition.

For this reason a thinner fluoroscopic screen is used in the image intensifier than is provided for ordinary fluoroscopy. On the other hand the reduction in the screen thickness cannot be carried too far, for a reason other than that connected with efficiency, as will be seen presently.

In the case of the image intensifier there are also other possible causes of blurring. These must of course not detract from the available gain in definition, and in fact, in our model, they do not.

through which the screen is viewed. No appreciable blurring occurs in the photocathode either, since this is in direct contact with the fluoroscopic screen, and, in order to avoid too much absorption of the liberated photo-electrons, is extremely thin.

Owing to these favourable properties, the over-all blurring effect produced by the image intensifier is not more than 0.4 mm. In order that the advantage of this low "intrinsic" blurring may not be lost, the previously mentioned geometrical blurring must also be limited as much as possible. It is therefore desirable to use in conjunction with the intensifier an X-ray tube having a very small focus (fine focus tube).

Contrast

On passing through the subject to be examined, the X-rays are attenuated to an extent that varies from one point to another, and it is the resultant "radiation contrast" that supplies the required information regarding the subject, i.e. the image. The processes involved in the conversion of the X-rays to fluorescent light, then into photoelectrons and finally into light again, are wholly linear, owing to the very small electronic current occurring in the image intensifier. In principle, therefore, the contrasts in the image are unchanged ($\gamma = 1$), in contradistinction to, for example, photographic reproduction (in which case γ may be as much as 2.5 to 3).

For various reasons, however, the image intensifier does occasion a certain amount of fogging, i.e. a brightness contribution more or less uniformly distributed over the whole image, which results in a loss of contrast. The glass window of the tube through which the X-rays enter partly scatters the rays and thus delivers diffuse radiation to the fluoroscopic screen; similarly those X-rays which are not absorbed by the fluoroscopic screen but which fall on the wall of the tube are in part re-diffused towards the screen. Again, a fraction of the blue fluorescent light which the very thin photocathode is not capable of absorbing is reflected from the inner wall of the tube, back to the photocathode, where it liberates electrons in the wrong places; up to a point this also applies to the persistence of the fluoroscopic screen. Furthermore, the photocathode exhibits a certain amount of thermionic emission which is independent of the intensity of illumination received by it; field-emission is also liable to occur at the photocathode. Both these effects contribute additional light uniformly spread over the whole of the viewing screen. A proportion of the total number of electrons arriving at the viewing screen is reflected there and tends to liberate positive ions from the walls of the anode space (as well as from possible residual gases), and these ions may be diffused in the direction of the anode aperture where they are accelerated by the field in the direction of the photocathode; here they may liberate more electrons and this will also produce additional diffuse light. Finally, at the viewing screen there is also some diffusion of the light generated by the screen itself, resulting in a kind of halo which is intensified by repeated reflection in the glass support of the screen, whose granules are in contact with it.

Very severe fogging would occur if a part of the light from the viewing screen were to return to the

photocathode (a kind of "feedback"), but this is prevented by coating the rear of the viewing screen with a thin layer of aluminium which transmits practically all the high-velocity electrons, but completely intercepts light that would be reflected back¹⁰).

Suitable precautions have also been taken to reduce to a minimum all the other sources of fogging summarised above. As a result the loss in contrast is kept within reasonable limits. We will mention only one or two of these precautions. For example, the inner lining of the tube is of a kind that does not reflect blue light too well. The shape of the electrodes is such that high field strengths do not occur at the photocathode, and field emission is accordingly low. For the same purpose and also in order to limit thermionic emission, the threshold potential of the cathode should be fairly high.

Accurate data regarding the degree of fogging cannot be given at this time, but for orientation purposes we may assume that this represents 10% of the total luminous flux which the viewing screen would emit without the fogging. (This value is probably not very far from the actual figure.) From the aspect of contrast this means that where we would have had to observe contrast between a brightness of 105 and an average of 100 for example, we will now have a contrast of 115:110. The loss is thus of little importance. When contrast is very pronounced the depreciation is greater (10 as against 100 would become 20 to 110), but such contrasts are generally easily distinguished in any case.

Principle limitation of minimum perceptible subject detail

Let us now ascertain in how far the radiologist, using the image intensifier, really will be able to perceive more detail, that is, details of smaller dimensions and lower contrasts, than with the ordinary fluoroscopic screen. The ultimate answer to this question can be obtained only with the intensifier in actual use. While, however, insufficient medical experience has been gained as yet with the instrument to enable one to express in numerical values the advance which the introduction of the intensifier undoubtedly represents, it is possible to show theoretically that the minimum detail to be perceived cannot decrease below a certain limit. This is attributable to the circumstance that X-rays are in the form of quanta; attention was first drawn to this consequence by the scientists Sturm and Morgan¹¹).

The energy quanta, or photons, of magnitude

¹⁰ See J. de Gier, Philips Techn. Rev. 10, 97-104, 1948.

¹¹ R. E. Sturm and R. H. Morgan, Screen intensification systems and their limitation, Amer. J. Röntgenology Rad. Ther. 62, 617-634, 1949.

$h\nu$, of which all radiations consist ($h = \text{Planck's constant}$ and $\nu = \text{frequency of oscillation}$), are not emitted by the source perfectly regularly, nor are they absorbed regularly by a medium such as the fluoroscopic screen or the retina of the eye. It is possible to speak of an average number \bar{N} of quanta per given interval of time, on a given surface, and this determines the intensity of the radiation emitted or absorbed. In the case of a number of such time intervals, however, the number of quanta within each interval will not be equal to \bar{N} ; random fluctuations about this average occur, whose order of magnitude, computed on the basis of probability, is given by the so-called standard deviation $\sqrt{\bar{N}}$. The order of magnitude of the relative fluctuations in the radiation is defined by the coefficient of variation, which is equal to $\sqrt{\bar{N}}/\bar{N} = 1/\sqrt{\bar{N}}$.

These natural fluctuations, the "noise" in the radiation, set a fundamental limitation on the detail which the radiation is capable of conveying, or, in more concrete terms, on the minimum radiation contrasts which can be communicated by means of the quanta absorbed.

This argument is immediately applicable to the human eye. Let us take as starting point a surface element of a given object, with a size of 0.1 mm^2 . Those quanta will contribute towards the perception of brightness which travel from this surface element to the pupil of the eye within the "accumulation time" T of the eye. Only a fraction of these is effectively absorbed in the retina, i.e. produce the stimulus conveyed to the brain by the optic nerve. The fluctuations in this number of quanta can be regarded as fluctuations, in effect, in the observed brightness of the surface in question, and it is found that the difference in brightness between this surface and its background definitely cannot be perceived if the contrast is not at least 3 to 5 times as great as the relative fluctuation already mentioned. The lower limit of perceptible contrast is thus expressed in terms of the "noise" in the light quanta effectively absorbed. It will now be seen at once that as the size of the surface observed is increased (i.e. more quanta are permitted to contribute to the sensation of brightness), so also does it become possible to appreciate lower contrasts; see for instance the well-known contrast-detail diagrams obtained in the fluoroscopic examination of phantoms (artificial test-objects), fig. 5¹²⁾. It will also be clear that when the brightness level is raised, contrast sensitivity increases.

However, this takes place more slowly than might be concluded from the decrease in the relative fluctuations, since, when the brightness increases, the percentage of light quanta effectively absorbed drops (owing amongst other things to a saturation effect in the optic nerve), as also does the accumulation time T .

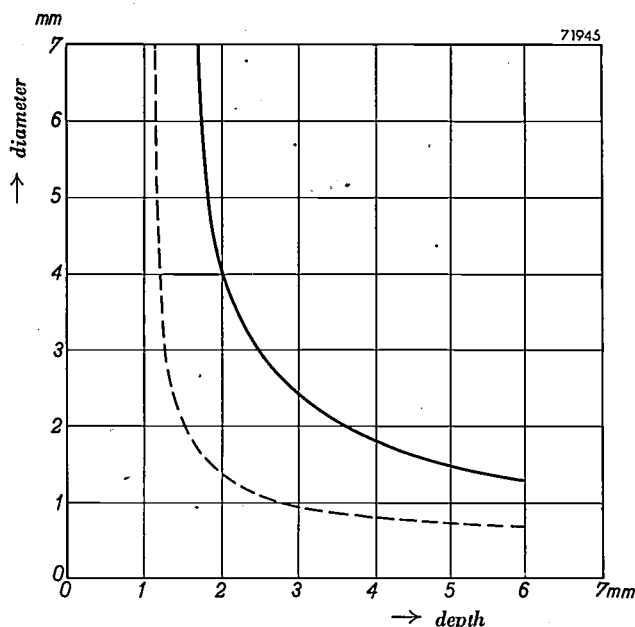


Fig. 5. Example of a contrast-detail diagram as obtained with the aid of an X-ray "phantom". The "phantom" consists of a number of plates of "Philite", one of which has in it a number of holes of different diameters and depths. The full line represents as a function of the hole diameter, the depth of a hole (contrast!) necessary to make it just visible in normal fluoroscopy. (The dotted line shows the results of provisional tests with the image intensifier under the same conditions.)

Let us now turn to X-ray diagnostics, again taking a surface of 0.1 sq. mm of the fluoroscopic screen. In chest examinations 100 000 X-ray quanta fall on such an area per second. Of this the fluoroscopic screen absorbs 65%. The accumulation time T of the eye in normal fluoroscopy is about 0.1 sec. The number of X-ray quanta involved in providing information about details of the above-mentioned size of the lungs under examination is therefore about 6500. One X-ray quantum of 70 kV absorbed liberates an average of 5000 light quanta in the screen. Thus we have more than 3×10^7 effective quanta; these occur in "packets" of 5000 quanta at a time, however. Consequently — and this is the cardinal point of the whole argument — the large number of quanta that would yield the low relative fluctuation of $1/5600$ (if all the quanta occurred individually) cannot furnish any more information than the original 6500 X-ray quanta were capable of doing, with the much greater fluctuation of $1/80$.

¹²⁾ See G. C. E. Burger, Phantom tests with X-rays, Philips Techn. Rev. 11, 296, 1950.

In normal fluoroscopy this is not really important. In fact, of the light quanta obtained, only a very small proportion falls on the pupil of the eye (see fig. 6): in the retina only 120 quanta are

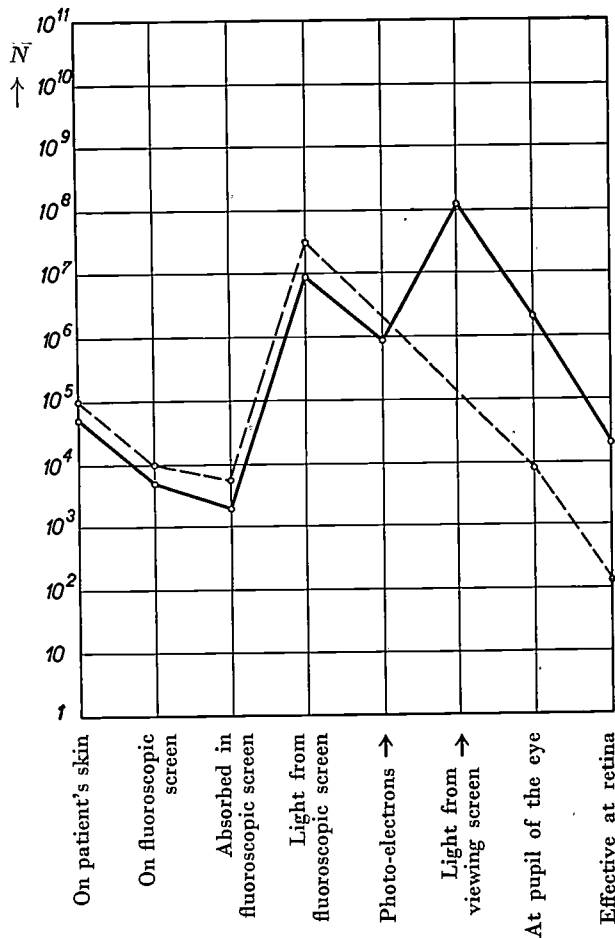


Fig. 6. Number of quanta occurring at different stages during a chest examination by means of X-rays, within a beam corresponding to an area of 0.1 sq. mm on the fluoroscopic screen. Voltage on X-ray tube 70 kV, current 2.5 mA, distance from X-ray focus to screen 70 cm.

Dotted line: ordinary fluoroscopy; accumulation time 0.1 sec.

Full line: with image intensifier; the accumulation time is now slightly shorter.

effectively absorbed, which is much less than the available number of X-ray quanta. In X-ray images of objects 0.1 sq. mm in size, therefore, the eye will not perceive contrasts of less than 3 to 5 times 1/11 and a large part of the information contained in the X-ray quanta is thus not used.

The image intensifier, however, has entirely altered this state of affairs. Initially the situation is rather less favourable, since, out of the above mentioned 100 000 X-ray quanta falling per second on 0.1 sq. mm of the thinner fluoroscopic screen, fewer are absorbed by the thinner screen and, moreover, the accumulation time *T* of the eye is shorter at the higher brightness level of the intensifier. We may take the number of effectively

absorbed quanta to be about 1750. The light quanta produced by them in the fluoroscopic screen, however, are now put to better use (see fig. 6); it may be estimated that the area observed now yields roughly 20 000 actively absorbed light quanta, which is considerably more than the number of active X-ray quanta. This means that contrast sensitivity is then high enough for the eye to register the whole of the information contained in the X-ray quanta — but, obviously, not more.

This bears a direct analogy with optical enlargements. For example, in the image produced by a microscope objective, the fine details of the object are resolved only up to a certain limit; the "information" regarding the object is limited by the resolving power of the objective (which is in turn determined by the lens aberrations and ultimately by the wavelength of the light). The resolving power of the eye is much lower and the image is therefore observed through a high-magnification eye-piece. Now the greatest effective magnification of the eye-piece is that at which the details as separated by the objective can be seen as separate by the eye. If the eye-piece is any stronger, the details are certainly seen at larger angles, but the wealth of detail is not increased, the information contained in the image being by then already exhausted. It may be said that the extra enlargement is "ineffective". Similarly it may be said of the image intensifier that the part of the brightness intensification which corresponds to the factor by which the number of effective light quanta in the eye exceeds the number of effective X-ray quanta in the fluoroscopic screen, may be regarded as ineffective.

The conclusions to be drawn from the above argument are as follows:

- 1) Of the brightness intensification factor of 1000 which our instrument yields, a maximum factor of about 20 is "effective" ¹³⁾; the remaining factor of 50 plays no part in reducing the smallest perceptible contrast, but is of great practical significance as it makes vision easier (no adaptation necessary).
- 2) Should it be found desirable to make use of the properties of the image intensifier to reduce the X-ray dosage to the patient (for the same duration of observation), this will be unavoidably

¹³⁾ This factor is greater than the ratio 1750/120 to be derived from the example in fig. 6 since, owing to the decrease in the accumulation time *T*, as also in the percentage of quanta producing stimuli in the optic nerve, a relatively higher brightness is required to ensure that the eye absorbs as many effective light quanta as there are effectively absorbed X-ray quanta.

accompanied by a loss of "information" which is, after all, determined by the number of X-ray quanta. If the same amount of information as that obtained from the ordinary fluoroscopic screen is considered adequate, the dosage for the purpose of a chest examination could, theoretically, be reduced at most to about 1/15th; this would still leave in hand a gain in brightness to a factor of about 60, which would represent a marked decrease in adaptation time. For thicker subjects the permissible reduction factor would be smaller.

The image intensifier and radiography

Similar considerations to those outlined above will make it clear that the use of the image intensifier can never provide the same high quality of image as that obtained in X-ray photography (radiography). The photograph owes its superiority to the fact that the accumulation time of the photographic emulsion is not limited, as is that of the human eye, added to which the emulsion retains what it has accumulated, the dosage time being thus no longer than the accumulation time. Unless a series of exposures is to be made, the dosage is therefore not usually dangerous and the intensity of the X-rays and/or the exposure time can be adjusted for each subject so as to yield the best average density. For a representative grade of film it will be found that to this end 20 000 X-ray quanta per 0.1 sq.mm (about 0.001 ergs of energy) must be effectively absorbed. The exposed film is examined with the aid of a viewing lantern; the illumination level can be raised as desired and matters so arranged that the light quanta received by the retina are in any case sufficient for perception of all the information the photograph has to offer (fig. 7). Now the difference in image quality between normal fluoroscopy, image intensifier and radiography is immediately expressed by the difference in number of useful quanta, i.e. for a chest examination 120 : 1750 : 20 000. With subjects that entail a higher degree of absorption (e.g. gastric fluoroscopy), the supremacy of the photograph is still more pronounced; it remains at 20 000 quanta, whereas in the two other cases the numbers of quanta are much lower.

Under these circumstances it will be obvious that the principle of the electronic image intensifier has no advantage over conventional X-ray photography. It can be employed to advantage, however, in fluorography. In this Review, on several occasions, the fact has been discussed that in this technique,

whereby a reduced X-ray image is obtained with the aid of a camera, it is a difficult matter to obtain sufficient light to expose the film, and this is the reason why extremely fast lenses have been developed for this class of work¹⁴⁾. The image intensifier, which furnishes a very bright, reduced image that is capable of being photographed quite easily on a 1 : 1 scale, provides a very convenient solution to the problem. Apart from this, the image intensifier may possibly prove to be the answer to the old problem of X-ray cinematography. As far as the mechanical aspect is concerned, X-ray fluorography has already brought a solution within sight, but the next step towards successful X-ray cinematography, although apparently only a small one, calls for some caution, since a very large number of photographs taken in succession means that the total X-ray dosage to the patient will again be fairly high. A reduction in the X-ray intensity with a view to limiting the total dosage

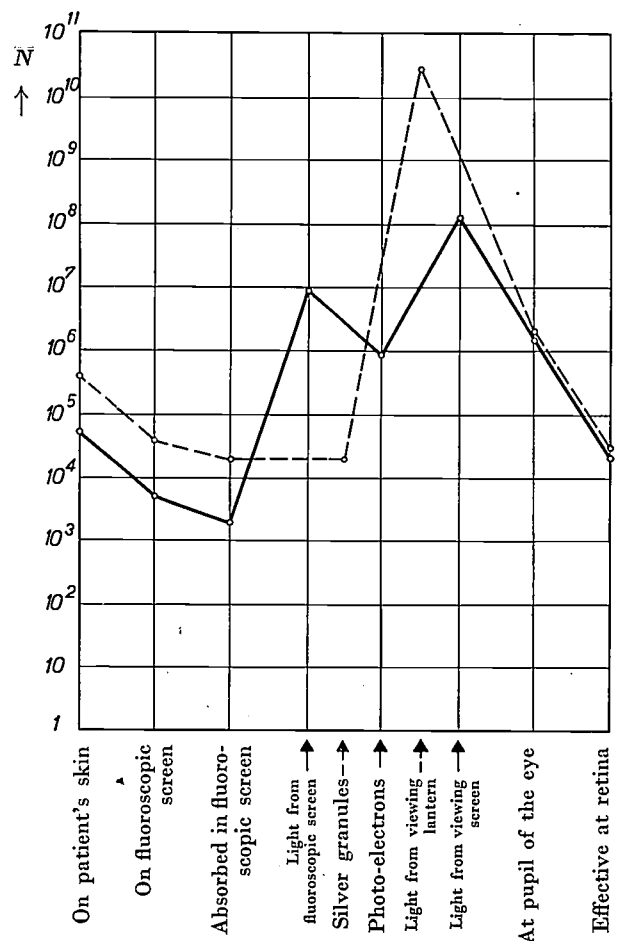


Fig. 7. Same as in fig. 6 but for normal radiography (dotted line). Full line; with image intensifier, as reproduced from fig. 6.

¹⁴⁾ W. Hondius Boldingh, Fluorography with the aid of a mirror system, Philips Techn. Rev. 13, 269-281, 1952 (No. 10).

during the taking of the film is possible only up to a certain point, seeing that even with very fast lenses too little light is otherwise obtained to ensure sufficient density of the film. This difficulty is overcome by the extra brightness provided by the image intensifier; even at appreciably lower X-ray intensities than those usually employed in fluorography, the image on the viewing screen will ensure a well-exposed film without necessitating particularly fast lenses. The fact that a reduction in the X-ray intensity detracts from the image quality (available information) has already been explained. It is a well-known phenomenon, however, that some slight sacrifice in image quality can safely be made in cinematography, since the summation of the impressions received from the successive pictures provides additional information.

Summary. In order to reduce to a minimum the risk of injury to the patient as a result of X-ray examination, the X-ray dosage that the patient receives must be limited as far as is practicable. Therefore, the attainable brightness of the image on a standard fluoroscopic screen is so low that long visual

adaptation periods are required before such images can be studied; even then, visual acuity and the contrast sensitivity of the eye are both fairly low at such low brightness levels. Improvement can be brought by increasing the brightness of the fluoroscopic image ("intensifying"). The electronic tube by means of which this is now made possible functions on the principle of the image converter. The X-rays produce on a fluoroscopic screen an image whose radiation liberates electrons from a photocathode in contact with the screen. The electrons are focussed by an electrostatic electronic lens system working at a potential of 25 kV, to produce another image, reduced 9 times in size, on a second fluorescent screen, the viewing screen, which is observed through a simple microscope of $9 \times$ magnification. The gain in brightness, to a factor of 800-1200 with respect to normal fluoroscopy, is produced partly by the accelerating field which imparts energy to the electrons, and partly by the electron-optical reduction of the fluoroscopic image. The resolving power of the image intensifier is higher than that of the best fluoroscopic screen. Owing to a number of causes the ultimate image is subject to a certain amount of fogging, which does reduce the contrast, although not to a serious degree. From a discussion of the lower limit of perceptible detail as based on spontaneous fluctuation in the number of X-ray and light quanta it is seen that the brightness can be intensified up to a factor of about 20 with a corresponding increase in the amount of "information" to be derived from the image. Beyond that point the information is restricted by the number of contributing X-ray quanta, but the much higher gain factor of the image intensifier offers the advantage that the image can be viewed without preliminary dark adaptation of the eyes, in a room with practically normal illumination. In conjunction with film cameras the intensifier paves the way to X-ray cinematography without the necessity for increasing the X-ray dosage to a hazardous extent.

MAGNETRONS

by J. VERWEEL.

621.385.16:621.396.615.141.2

Split-anode magnetrons have already been in use for a number of years for generating decimetric and centimetric waves. In 1937 a radio telephone link between Eindhoven and Nijmegen (Holland) employing magnetrons and operating on a wavelength of 25 cm was established. During the period 1940-1945 great strides were made in the development of magnetrons; output power was increased by more than 1000 times and generation of waves of 1 centimetre became a practical possibility. In this way the magnetron came to assume the main role in radar, and that this has had an important effect on war-time as well as peace-time activities is sufficiently well-known. In this article the working and construction of modern magnetrons are discussed.

Some 30 years ago Hull¹⁾ published an article on the effect of an uniform axial magnetic field on the motion of electrons between coaxial cylinders (fig. 1a). Although Hull spoke neither of oscillations, nor used the word magnetron, his work laid the foundations for the use of the magnetron as an oscillator. Since 1921 a varying but none the less steadily increasing amount of interest was shown in the magnetron. Up to 1930 the oscillations produced by these tubes were of little practical value, but from that time onwards there was a widespread increase in interest, as it was found that split anodes (fig. 1b) with resonators or transmission lines between them were capable of generating frequencies with a high degree of efficiency. Posthumus²⁾ has contributed much towards the development of valves designed on this principle and his theory of the oscillations produced by them is regarded as of pioneer value. In the four years following his publication more than 100 articles were written on the subject of the magnetron, in all parts of the world.

In the period up to 1940 interest waned slightly, but in that year Boot and Randall³⁾ carried out experiments with magnetrons provided with cavity resonators (fig. 1c), and these experiments led to the successful use of the magnetron as a transmitting tube for radar. Feverish activity ensued; output power was increased once more and some very short waves were produced. Important theoretical work was also submitted. To give some idea of what has been achieved with these tubes, we may mention

that in recent years magnetrons have been made which operate at a wavelength of 3 cm, delivering more than 1000 kW peak output, and that others operate at a wavelength of 3 mm, supplying an output of several kilowatts⁴⁾. This power is transmitted in pulses varying in duration from a few tenths of a microsecond to several microseconds, the repetition frequency being usually between 500 c/s and 4000 c/s. Fig. 2 illustrates a modern magnetron for 8.5 cm waves, delivering a power of 400 kW peak.

By far the most important field of application of the magnetron to date is radar, after which comes diathermy, where centimetric waves are in some cases particularly beneficial. Another special field is to be found in the linear accelerator, an appliance for imparting enormous velocities to electrons for the purposes of nuclear physical research⁵⁾.

In 1939 this Review contained an outline of the development of magnetrons up to that time⁶⁾; the present article is intended to provide a general review of modern magnetrons which will at the same time pave the way for later articles dealing with special designs in this sphere. In the present paper we shall endeavour to answer the following questions.

- 1) In what manner is the energy derived by the electrons in the magnetron from the direct voltage source converted to H.F. energy?
- 2) To what conditions must the resonators conform?
- 3) What kind of output system can be used to draw from the magnetron the energy developed by it?

¹⁾ A. W. Hull, Phys. Rev. 18, 31-57, 1921.

²⁾ K. Posthumus, Principles underlying the generation of oscillations by means of a split anode magnetron, Philips Transm. News 1, 11-25, Dec. 1934; Oscillations in a split anode magnetron, Wireless Eng. 12, 126-132, 1935.

³⁾ H. A. H. Boot and J. T. Randall, The cavity magnetron, J. Inst. Elec. Engrs. 93, 3a, 928-938, 1946.

⁴⁾ J. R. Pierce, Millimeter waves, Electronics 24, 66-69, 1951 (No. 1)

⁵⁾ Philips Techn. Rev. 14, 1-12, 1952 (No. 1).

⁶⁾ G. Heller, Philips Techn. Rev. 4, 189-197, 1939.

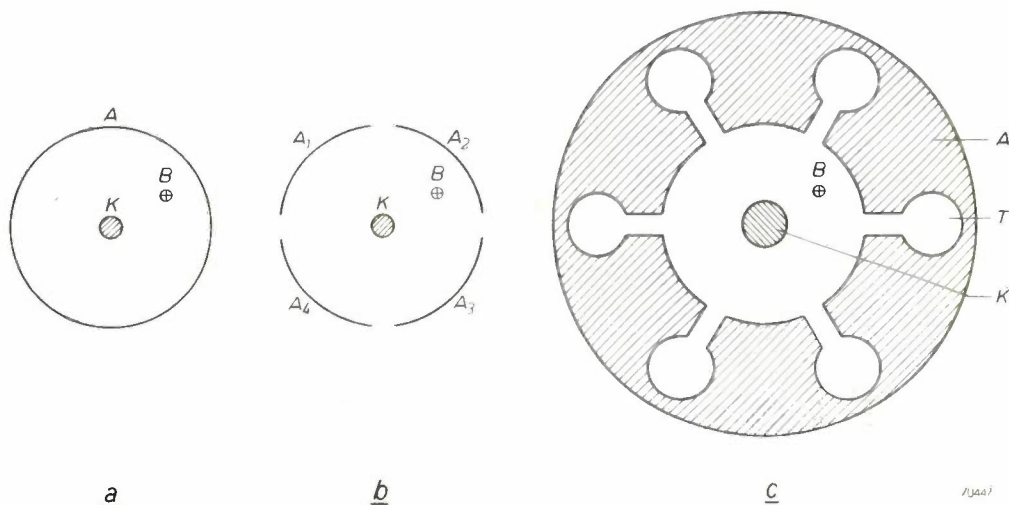


Fig. 1. Various forms of magnetrons in cross section. *K* = cathode; *A* = anode. The flux density *B* is perpendicular to the plane of the drawing and is uniform within the whole of the anode space. *a*) Hull magnetron. *b*) Magnetron with split anode, comprising four segments (*A*₁, *A*₂, *A*₃ and *A*₄). Such magnetrons can have a very high efficiency. *c*) Magnetron by Boot and Randall. The anode is a solid block of copper with cavity resonators *T* cut in it.

This will be followed by a few details regarding construction.

It will be necessary to restrict ourselves to the main essentials and to refer the reader to existing publications ⁷⁾ for a more detailed study.

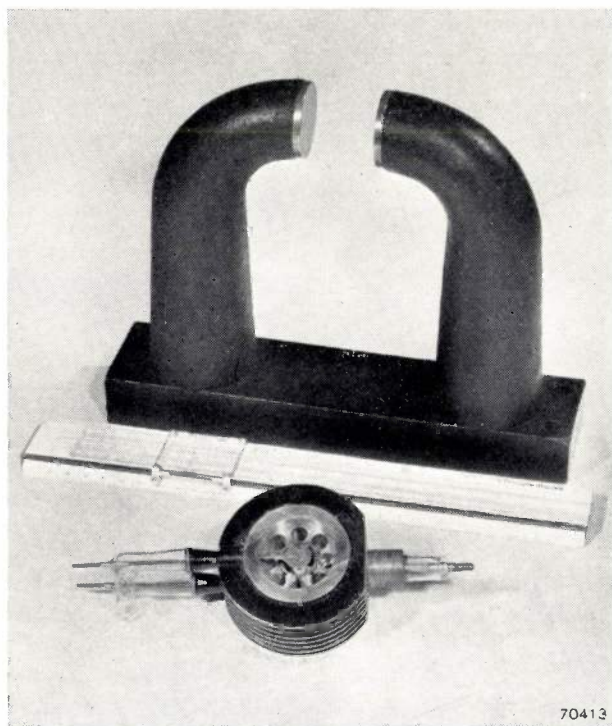


Fig. 2. Experimental magnetron for a wavelength of 8.5 cm, 400 kW peak output, with one of the covers removed. Behind the magnetron will be seen the associated permanent magnet.

⁷⁾ See e.g. G. B. Collins, *Microwave magnetrons*, Radiation Lab. Series 6, McGraw Hill, New York 1947, and J. B. Fisk, H. D. Hagstrum and P. L. Hartman, *The magnetron as a generator of centimeter waves*, Bell Syst. Techn. J. **25**, 167-348, 1946.

Motion of an electron in mutually perpendicular electric and magnetic fields

In a cylindrical magnetron the electric field is mainly perpendicular to the magnetic field. When the magnetron is oscillating, the electric field has not only a radial, but also a tangential component and, moreover, the space charge of the electrons has a very marked effect on the motion; in consequence the theory is rather complicated. As the various computations that have been made are very lengthy, they fall outside the scope of this article and we shall therefore confine ourselves to a general outline of the electronic motion.

Let us first call to mind the theory relating to the motion of electrons in a uniform magnetic field, the flux density of which is denoted by *B* (in the *z*-direction of a rectangular system of co-ordinates), and a uniform electric field *E* perpendicular thereto (in the negative *y*-direction). We are then concerned with the electric and the Lorentz forces. Analytically this is represented in our system of co-ordinates by:

$$\begin{aligned} m \, dv_x/dt &= ev_y B, \\ m \, dv_y/dt &= eE - ev_x B, \\ m \, dv_z/dt &= 0, \end{aligned}$$

where *m* is the mass of the electron, *t* is the time and *v_x*, *v_y* and *v_z* are the velocity components along the axis.

It is easy to verify the general solution of these expressions, viz:

$$v_x = -a \cos \frac{eB}{m} (t - t_0) + \frac{E}{B},$$

$$v_y = a \sin \frac{eB}{m} (t - t_0),$$

$$v_z = b.$$

a , b and t_0 are integration constants. Further integration gives us the equations for the electron paths. If we assume that all co-ordinates and velocities are zero at $t = 0$, we have:

$$\left. \begin{aligned} x &= \frac{E}{B} t - \frac{m}{e} \frac{E}{B^2} \sin \frac{eB}{m} t, \\ y &= \frac{m}{e} \frac{E}{B^2} \left(1 - \cos \frac{eB}{m} t \right), \\ z &= 0. \end{aligned} \right\} \dots \dots (1)$$

These equations represent a cycloid, as described by a point on the periphery of a circle rolling along the x -axis with constant velocity E/B (fig. 3). The diameter of the circle is $(2m/e) (E/B^2)$, and this represents the maximum distance of the electron from the x -axis. The general movement of the electron is perpendicular to the lines of electric and magnetic force (i.e. in an equipotential surface of the electric field) and the electron thus describes arcs to a height of $(2m/e) (E/B^2)$.

Now let us take the case of a plane magnetron, the cathode of which is formed by the plane $y = 0$ and the anode by $y = d$ (fig. 4). As long as the magnetron does not oscillate and, disregarding the

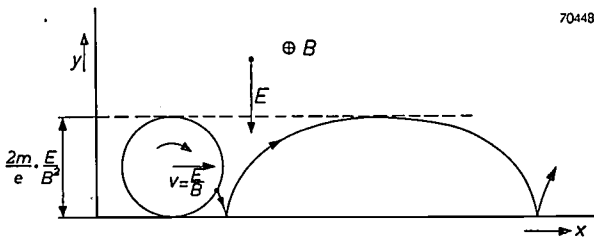


Fig. 3. Under the combined influence of a uniform electric field E and a perpendicular uniform magnetic field (flux density B), an electron emitted with zero velocity at the x -axis follows a cycloidal path. The average velocity in the x -direction is E/B .

space charge, the field between the cathode the anode will indeed be uniform. If, as assumed, the electrons leave the cathode with zero velocity, they will move along the cycloidal paths mentioned previously. If $(2m/e) (E/B^2) > d$, they describe only the first part of the cycloid shown in fig. 3 and are then absorbed by the anode. As soon as $(2m/e) (E/B^2) < d$, however, the cycloidal path is completed and the electrons do not reach the anode. The magnetic field whereby the path just touches the anode is known as the critical field, the flux density of which

is denoted by B_{cr} . Hence:

$$B_{cr}^2 = \frac{2m}{e} \frac{E}{d},$$

$$B_{cr} = \sqrt{\frac{2m}{e} \frac{\sqrt{V}}{d}},$$

where V is the anode voltage.

For a cylindrical magnetron, in which perfectly cycloidal paths are obviously no longer possible, a similar equation applies, viz:

$$B_{cr} = \sqrt{\frac{8m}{e} \frac{\sqrt{V}}{r_a (1 - r_k^2/r_a^2)}}, \dots (2)$$

where r_a is the radius of the anode and r_k that of the cathode. In a non-oscillating magnetron the anode current is therefore zero when $B > B_{cr}$.

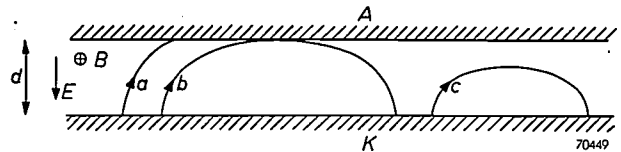


Fig. 4. Electron paths in a non-oscillating magnetron with distance d between cathode K and anode A (both plane):
 a) $(2m/e) (E/B^2) > d$: paths end at the anode;
 b) $(2m/e) (E/B^2) = d$: paths touch the anode;
 c) $(2m/e) (E/B^2) < d$: electrons no longer able to reach the anode.

When the electric field is no longer uniform, but nevertheless still perpendicular to the magnetic field, as in an oscillating magnetron, the motion is more complex. Assuming the magnetic field to be strong enough, so that the radius of curvature of the cycloidal path is small compared with that of the equipotential lines, the electric field in small areas can be regarded as uniform. In that case, too, the electron, on an average, follows an equipotential line. It is true that in the actual magnetron the radius of the circular motion is not usually small in comparison with that of the equipotential lines, but the approximation is close enough to justify a qualitative representation of the mechanism.

Qualitative explanation of the occurrence of oscillations

In the event that $B < B_{cr}$, the magnetron behaves as an ordinary diode and oscillation is not possible. It will therefore be investigated what happens when $B > B_{cr}$. In the absence of oscillation no anode current flows and the electrons follow cycloidal paths, their general movement being parallel to the cathode and anode. When oscillation

occurs the conditions are different; the electrons are then able to follow paths which do terminate at the anode; anode current flows and the magnetron consumes energy. Let us now assume that the condition of oscillation has been established in one way or another, and investigate how such oscillations are sustained.

Let the number of anode segments be $2n$ and assume the potentials between two adjacent segments to be constantly varying in counterphase, as is usually the case. Fig. 5 shows the lines of electric force of the alternating field at a given moment, in the event of four segments. According to Posthumus, this stationary alternating field is equivalent to fields rotating in opposite directions; the two fundamental components of these fields rotate with an angular velocity ω/n or $-\omega/n$, where ω is the angular frequency of the oscillation. Now, it is essential to the mechanism that the electrons be synchronised with one of these fields. The mean tangential velocity of the electrons, as we have seen, is E/B . An approximate expression for the synchronisation can be based on the velocity of the electrons half-way between anode and cathode, disregarding the fact that the electric field is dependent on the radius r . The angular velocity is then:

$$\frac{v}{\frac{1}{2}(r_a+r_k)} = \frac{E/B}{\frac{1}{2}(r_a+r_k)} = \frac{V/(r_a-r_k)B}{\frac{1}{2}(r_a+r_k)} = \frac{2V}{(r_a^2-r_k^2)B}$$

Equating this with the angular velocity ω/n , we then have:

$$V = \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a} \right)^2 \right\} \dots (3)$$

We shall refer to this equation again presently, but assume for the moment that it is satisfied by a suitable selection of the parameters.

Let us now examine the interaction between the electron and the synchronous field, assuming that both rotate in, say, a clockwise direction. The other rotating field passes the electrons very quickly; as this is an alternating field it will have alternately an accelerating and a decelerating effect. The electrons will, therefore, perform small oscillations about the average path, but these will be disregarded. Let us first see how the electron behaves in an area where the tangential field has a decelerating effect (at *A* in fig. 6). As the condition of synchronisation has been met, the electron will remain in this zone. It is true that the alternating field is usually smaller than the steady field, but the effect may nevertheless not be ignored.

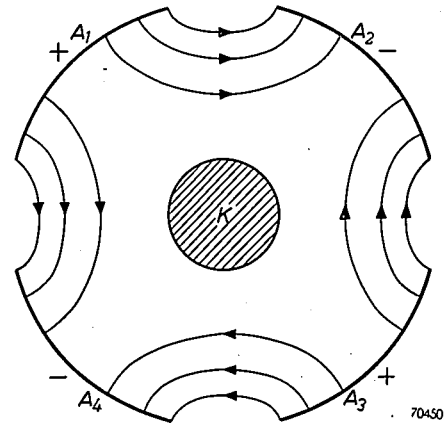


Fig. 5. The alternating electric field in a magnetron with four plates is a standing wave in the space between anode and cathode. This wave may be resolved into opposed rotating waves, or rotating fields. The electrons interact mainly with one of these fields.

The resultant E of the two thus forms a certain angle with the radial direction and, since the average motion of the electron is in the direction perpendicular to E , it will be seen that it is deflected towards, and ultimately reaches, the anode. Disregarding the fact that the electric field strength is dependent on the radius, the average velocity E/B of the electron, and therefore also the kinetic energy, will be constant. The potential energy dissipated by the electron in passing to the anode is thus not converted to kinetic energy as in the case of normal motion in an electric field, but is imparted to the alternating field as H. F. energy.

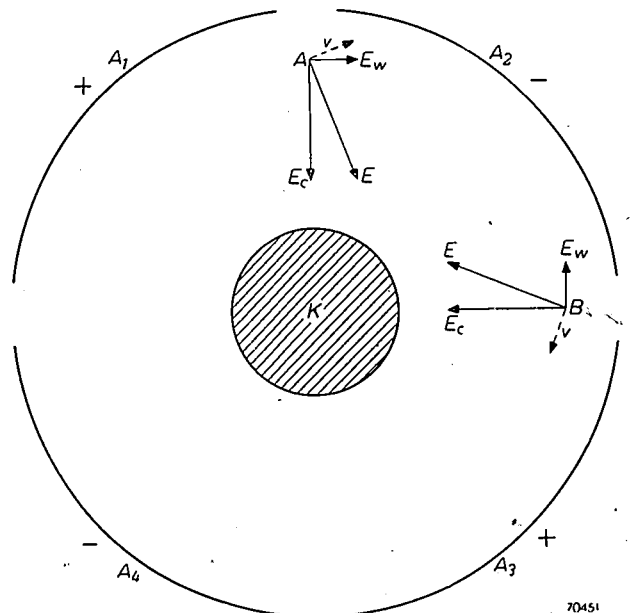
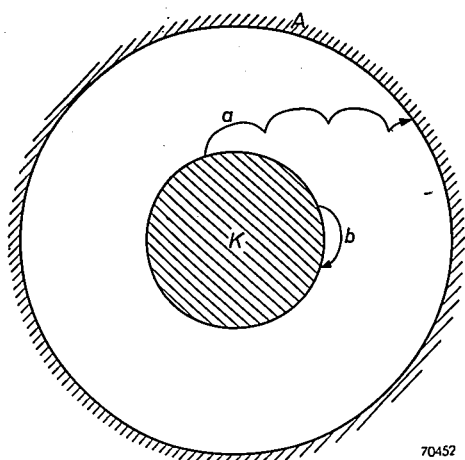


Fig. 6. The resultant E of the steady field E_c and the alternating field E_w , at two points, *A* and *B*. The mean electron velocity v is perpendicular to E . In a decelerating field E_w , v receives an outward-directed component, in an accelerating field E_w , an inward-directed component.

An electron in an area where a tangential accelerating field prevails (at *B* in fig. 6) will follow a path towards the cathode, since the field vector points in the other direction. The electron is accelerated by the H.F. field and is thrown back to the cathode after describing one arc of the cycloid. The paths of two such electrons are depicted in fig. 7; it will be seen that the first of the two describes a number of arcs (*a*), but that the second completes less than a single arc (*b*). The time required to describe one complete arc



70452

Fig. 7. Interaction of an electron with (*a*) a decelerating field and (*b*) an accelerating field. Electrons (*a*) are in the field longer than electrons (*b*), so that, averaging over all the electrons, energy is supplied to the field.

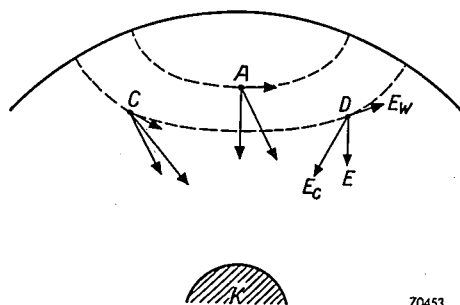
is constant and is, according to equation (1), equal to $(m/eB) \cdot 2\pi$. Seeing that the electrons *a* are in the field longer than the electrons *b* and therefore cede more energy than is absorbed by the electrons *b*, there remains a surplus, so that oscillation can be sustained and energy can be delivered to an external system.

It is, furthermore, a fortunate circumstance that electrons in a magnetron are driven towards a zone where a field of maximum deceleration exists and where the energy transfer is as favourable as possible; in effect, phase focusing occurs, as will be seen from fig. 8, in which the resultant *E* of the steady field *E_c* and the alternating field *E_w* are shown at the points *A* as well as at the points *C* and *D*, on either side of it. An electron situated at *C* for example, because it lags, will be in a stronger field than an electron at *A*. The mean velocity *E/B* of such an electron will thus be higher until the other field is overtaken. At *D* the electrons are in a weaker over-all field *E*; they therefore move more slowly until the field catches up with them.

In consequence of this, electron bunches occur at the most advantageous point and impart H.F. energy to the field.

At this juncture we may point out the difference that exists between a magnetron, and an "inverted" cyclotron. In a cyclotron the charged particles move under the influence of a magnetic field and are accelerated twice per revolution by a tangential electric field whose H.F. oscillations are synchronised with the circular motion. It would be possible to imagine this process inverted, that is to say, a process of shooting high-velocity electrons inwards; they would then yield energy to an H.F. field at the slots and would thus be decelerated. This is certainly possible in the magnetron, but the mechanism is rather different from that just described; if the oscillations were promoted by means of a suitable combination of potential and magnetic field, this would result in oscillations of very low efficiency. The angular frequency of the oscillations corresponds to that with which the cycloidal arcs are traversed in the magnetron: $\omega = eB/m$. In the oscillatory mechanism of the magnetron this frequency is of minor importance, however; in this case it is the mean velocity *E/B* that is important, where *E* contains particularly the constant-field component, which is entirely absent in the cyclotron.

The electronic motion can be computed quantitatively by iteration; the electron paths are calculated numerically for given values of the direct and alternating anode voltages and of the magnetic field. From a sufficiently large number of paths the space-charge distribution in the magnetron is then obtained, and from this follows a new potential field, which, in general, is different from the original. Further electron paths are next computed from this new field; these in turn give rise to a third field and so on. With a judicious choice of the original field, these successive approximations converge upon the correct solution. In 1928 Hartree⁸⁾ employed this method to quantum-mechanical path computations in atomic theory and, during the last war, the electronic paths in magnetrons



70453

Fig. 8. The resultant field vector *E* at a point *A*, and at two other points *C* and *D* on either side of *A*. Owing to the influence of the radial alternating field, *E* is greater at *C* and smaller at *D* than at *A*.

⁸⁾ D. R. Hartree, Proc. Cambridge Phil. Soc. 24, 111-132-1928.

were computed in this manner under his guidance⁹⁾.

The result of such an iteration procedure as applied to a magnetron having eight anode segments is seen in *fig. 9*. The paths of four electrons emitted by the cathode in different phases of the H.F. oscillation are depicted in a system of co-ordinates which rotates with the field. An elec-

co-ordinates, describes the same path within the system. The envelope embracing all these paths, shown by the broken line in *fig. 9*, represents the boundary of the space-charge cloud. This consists of four "spokes", each of which lies roughly at the location of the maximum deceleration field. They rotate together with an angular velocity $\omega/4$, and

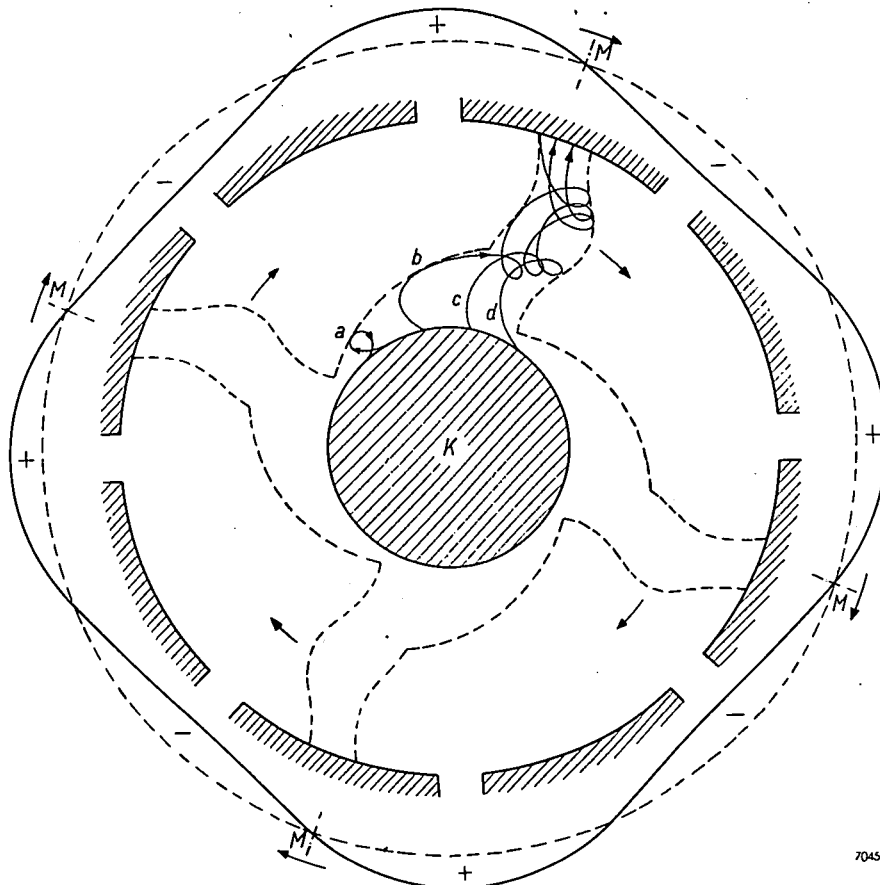


Fig. 9. Paths of four electrons (*a, b, c* and *d*) emitted by the cathode *K* in different phases with respect to the rotating field, plotted in a system of co-ordinates that rotates with the field. The actual field is constant with time. From a point on the cathode which rotates with the system of co-ordinates, the same path is traced each time. The envelope of all the paths is shown by the broken lines; this is the boundary of the space charge. The areas of the tangential field of maximum deceleration are shown by the lines *M*. The radial distance between the dotted circle and the curve fluctuating around it represents the rotating alternating field. (From an article by J. B. Fisk, H. D. Hagstrum and P. C. Hartman, *Bell Syst. Techn. J.* 25, 195, 1946).

tron *a* emitted at an unfavourable instant returns to the cathode; the three other, favourably emitted electrons *b, c* and *d* proceed to the anode and, in so doing describe a number of loops. Since the system of co-ordinates is rotating, the field is constant with time; hence every electron emerging from the cathode surface at a point which rotates with the

constantly pass those parts of the anode that are just about to become positive, intensifying the charge by reason of the induction. In this way the oscillation is sustained in much the same way as that of a pendulum to which a pulse is applied each time at the correct moment. It is thus clearly seen that the potential energy dissipated by the electron and not converted to kinetic energy, becomes available as H.F. energy.

⁹⁾ C. V. D. Report, Mag. 36 and 41.

Characteristics of the magnetron

In this section we shall trace the relationships that exist between current, voltage, magnetic field, flux density, etc.

We have already mentioned the cut-off condition indicating the smallest flux density necessary for oscillation, viz:

$$B_{cr} = \sqrt{\frac{8m}{e}} \frac{\sqrt{V}}{r_a(1-r_k^2/r_a^2)} \dots (2)$$

Further, we have derived the approximate condition for synchronism.

$$V = \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a} \right)^2 \right\} \dots (3)$$

A more accurately defined condition for synchronism (which will not be discussed here) is as follows:

$$V = \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a} \right)^2 \right\} - \frac{m}{2e} \left(\frac{\omega r_a}{n} \right)^2 \left\{ 1 - \left(\frac{r_k}{r_a} \right)^2 \right\} \dots (4)$$

Hartree formulated a condition that has to be met if the electrons are to reach the anode in the case of an infinitely small H.F. field. This states that:

$$V \geq \frac{\omega}{2n} r_a^2 B \left\{ 1 - \left(\frac{r_k}{r_a} \right)^2 \right\} - \frac{m}{2e} \left(\frac{\omega r_a}{n} \right)^2 \dots (5)$$

In *fig. 10* all these lines are shown diagrammat-

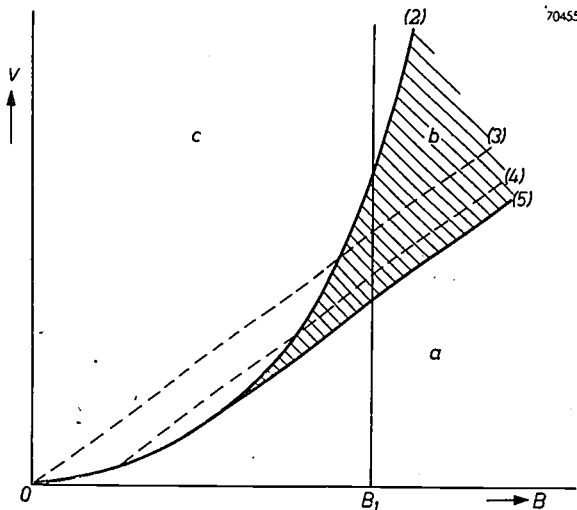


Fig. 10. The curve (2) representing the critical flux density and the Hartree line (5) divide the V - B plane into three zones a , b and c . In (a) the magnetron passes no current and no oscillation occurs. (b) is the oscillation zone; here the current rises steeply with the voltage. In (c) the magnetron is no longer cut off so that no oscillation occurs.

Lines (3) and (4) represent the simple and the more accurate conditions for synchronism respectively.

For the equations of lines (2), (3), (4) and (5) see the formulae of the same numbers.

ically, with the flux density B as abscissa and the voltage V as ordinate. The cut-off curve (2) is a parabola; the "simple" condition for synchronism (3) is a straight line through the origin, and Hartree's condition (5) produces another straight line, parallel to the other and tangent to the parabola. The more accurate condition for synchronism (4) is represented by a line located between the other two straight lines and also lying parallel to them.

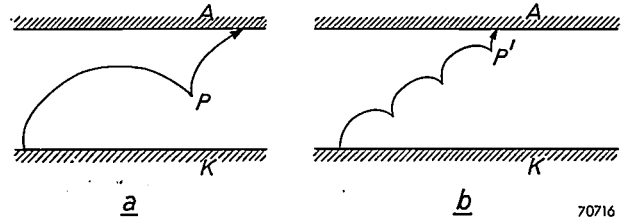


Fig. 11. Electron paths in a plane magnetron with cathode K and anode A .

a) B slightly higher than B_{cr} . The electron describes a large arc. As its velocity is zero at P , and as it can cede only little energy to the field on its way to the anode, a large proportion of the potential energy which it possesses at P is converted to heat. The efficiency is therefore only low.

b) B much greater than B_{cr} . The electron describes a small arc. Only that potential energy which the electron still possesses at P' can be converted to heat. The efficiency is now high.

Lines (2) and (5) divide the V - B plane into three zones a , b and c ; if the voltage be increased for a given value B_1 of the flux density, each of these zones will be traversed in turn. At low potentials zone a is involved; the magnetron is then cut off, but oscillation is not possible, for the voltage does not conform to condition (5). When the voltage is raised so that zone b is entered, oscillation can occur. With further increases in potential the anode current rises rapidly and, with it, the amplitude of the oscillation. Usually a limit is then encountered owing to the fact that the cathode is unable to deliver more current. The oscillations, which are at a maximum in the region of line (4), decrease when the voltage is raised, seeing that the condition for synchronism is then not met. In zone c current may flow without H.F. voltage, the magnetron is no longer cut off and no further oscillation occurs.

The efficiency of the oscillation is found to rise when the flux density is increased (hence with increasing voltage in order to remain on line (4) which represents the condition for synchronism). This will be readily appreciated from the figure. If the flux density B is but slightly higher than the critical value B_{cr} , the conditions will be as shown in *fig. 11a*. As is seen from equation (1), the height of the arc of the cycloid is not much less than the

distance from anode, to cathode. An electron leaving point P roughly at zero velocity still possesses much of its potential energy and can impart only a small proportion of it to the H.F. field; a large proportion is thus dissipated on collision with the anode. The efficiency is then relatively low.

However, if B is much greater than B_{cr} , we see from fig. 11b that the electron cannot lose more potential energy than it still possesses at the point P' and as this is much less, less is dissipated and the efficiency is higher.

A magnetron for a wavelength of 3 cm may in this way yield efficiencies of 50%, which is much higher than is the case with klystrons or travelling-wave tubes. Electron bunching takes place in magnetrons in the same way as in these other types of tube and, moreover, the bunching is synchronised with the wave in magnetrons as in travelling-wave tubes. The difference, however, is that in the magnetron the magnetic field has a phase focusing effect. As with an ordinary top, a force in the one direction, in this case tangential, is converted to a displacement in the perpendicular direction, i.e. radial. The velocity in the direction of the field remains the same; there is no dissolution of the bunches, but the potential energy lost by the electron on its way to the anode is imparted, in the form of H.F. energy, to the alternating field, as already stated above. Accordingly, almost the whole of the D.C. energy is effectively used.

The characteristics which are usually plotted for magnetrons are combined to form what is known as a performance chart; the anode current is plotted horizontally and the anode voltage vertically. By varying the voltage and the magnetic field the various points in the diagram are obtained. The output power and efficiency are measured, and lines of constant flux density, power and efficiency are drawn in after the manner of fig. 12. The different phenomena which we have just described then become apparent. The anode voltage is roughly proportional to the flux density and depends but little on the current strength. Efficiency is increased with the flux density, as already pointed out. Further, the efficiency is to a certain extent dependent on the anode current; with high values of the current, efficiency is impaired, since the space-charge then disturbs the phase focusing.

In the direction of higher voltage and current the performance of the magnetron is restricted by the cathode emission, disruptive discharges in the tube and by the maximum magnetizing force which the magnet is capable of producing. For normal use,

magnetrons are equipped with a permanent magnet (fig. 2), for which reason only those points in the performance chart can be attained which lie on a line of constant flux density.

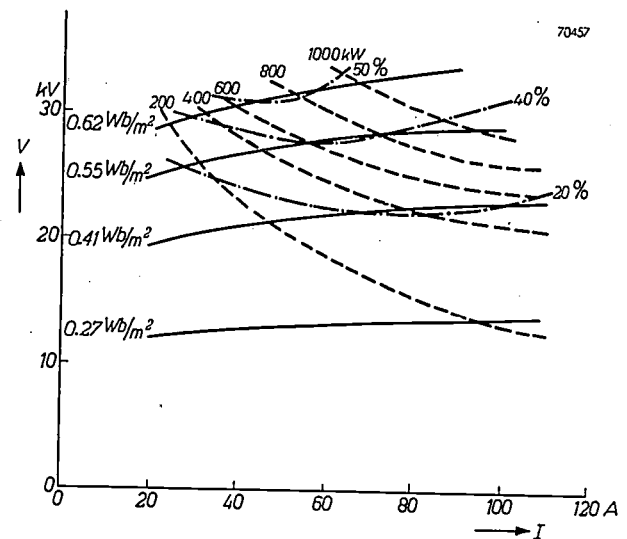


Fig. 12. Performance chart of a magnetron for a wavelength of 3 cm. The current is plotted horizontally and the voltage vertically. Lines of constant flux density (expressed in Wb/m²), constant power (in kW) and constant efficiency (%) are shown in the chart. This particular chart was prepared from a magnetron having 18 cavities; $r_k = 0.32$ cm; $r_a = 0.52$ cm; anode length = 2 cm.

Resonators

The H.F. energy occurs mainly in the resonators, where it oscillates in the form of electric and magnetic energy. The resonators determine the wavelength and also serve as reservoirs for the energy. A portion of the energy in the resonator is fed to the aerial or to another form of load.

Prior to 1940 the preferred form of resonator was a two-wire transmission line. In two-plate magnetrons the two anode sections were connected direct to the two wires (fig. 13a), whereas in magnetrons with four or more plates connection was established in pairs by means of the shortest possible leads, to keep the pairs at the same voltage; the transmission line was then connected to these leads (fig. 13b).

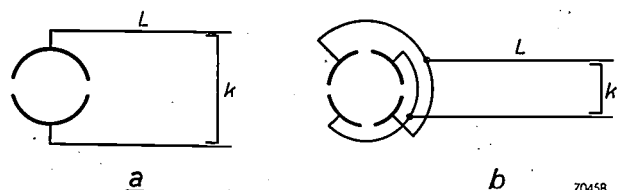


Fig. 13. a) Two-plate magnetron with transmission line L . Tuning is effected by means of the shorting bridge k . b) Four-plate magnetron with plates connected in pairs. The transmission line is welded to these connections.

Present-day magnetrons usually work with several resonators in the form of cavity resonators, cut out of a solid block of copper which serves as anode. The dissipative and radiation losses in such resonators are smaller, the dissipation of heat is greatly simplified and the resonators can be made more easily with a sufficiently high degree of accuracy for use on short wavelengths. Conventional forms of magnetron anode are depicted in fig. 14.

certain limit. In any case B must be greater than B_{cr} and accordingly cannot be reduced *ad libitum*. E must not be too high either, or we run into impractically high voltage values (50-100 kV). A fairly large number of cavities is therefore unavoidable.

As the cavities are mutually coupled via the end chambers and the central hole, the whole system has a series of natural frequencies, numbering $n + 1$ in a magnetron with $2n$ identical cavities.

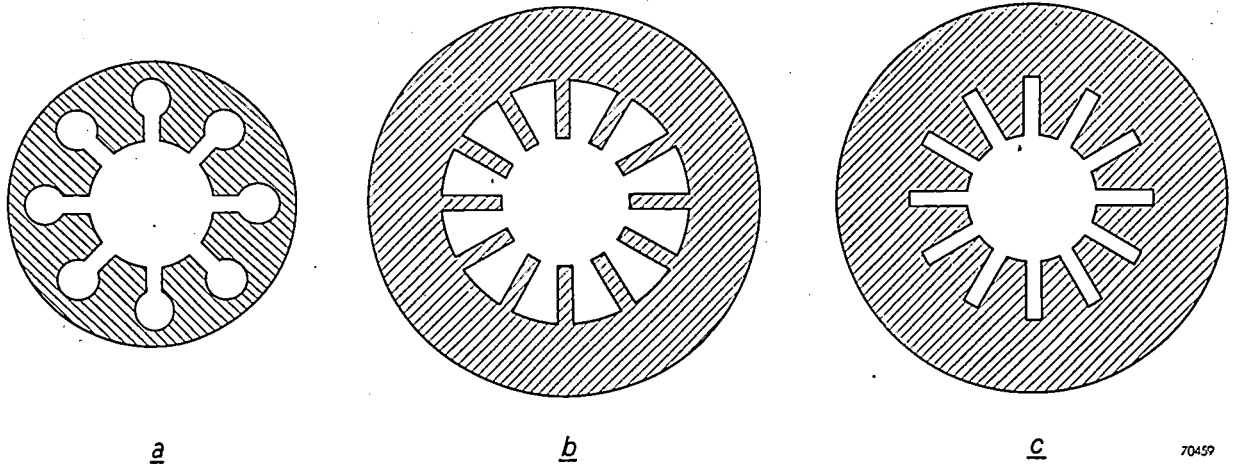


Fig. 14. Anode blocks for magnetrons (shown in cross section). The cavities consist of (a) holes and slots; (b) sectors; (c) slots.

In the case of resonators consisting of holes and slots it is still permissible to speak of lumped inductances and capacitances. The capacitance is concentrated particularly in the slots and we may also refer to a certain potential across the slots. The inductance may be localised in the holes through the walls of which the current flows. In other forms of resonator it is not possible to make this distinction, but it is still true that the electric field is concentrated mainly in the region of the open end, whereas the current is highest at the closed end, the magnetic field therefore being strongest in the latter area. (The fact that the capacitance and inductance cannot therefore be clearly distinguished, naturally does not imply that the natural frequency and impedance of the cavity resonator cannot formally be defined by an "effective" capacitance and inductance.)

The number of cavities in a modern magnetron is fairly large, being usually eight or more. In order to deliver the required high current strength, the cathode is given a large diameter. The diameter of the anode is then also large, but the spacing of the cavities, which is roughly equal to (E/B) times the half-cycle, cannot be increased beyond a

A system of two coupled circuits is able to oscillate in two different ways; the oscillations may be either in phase or in antiphase similarly, the modes of oscillation in a system of more than two coupled cavity resonators are distinguished by their mutual phase displacements. The phase displacement φ between two successive cavities is always constant; if there are $2n$ cavities, we find $2n$ times that phase difference over the cavities in turn, round to the starting point again, and this value must be a whole multiple of 2π :

$$2n\varphi = k \cdot 2\pi, \text{ or } \varphi = k \frac{\pi}{n},$$

where k is any whole number. We then speak of the mode number k . For every value of k there is a natural frequency; for $k = 0, 1, 2 \dots n$, there are $n + 1$ different natural frequencies. When $k = n$, $\varphi = \pi$ and this is usually referred to as the π -mode. Two adjacent cavities are then in antiphase. When $k > n$, other modes of oscillation occur, but the oscillation $n + m$ has the same natural frequency as $n - m$. Such modes of oscillation are known as degenerate. The occurrence of degenerate modes is due to the circular

symmetry of the anode block; the two field configurations are at a given moment images of each other with respect to a given plane through the axis. Under conditions of oscillation which are degenerate, that oscillation may easily be excited that is not coupled with the electronic motion, and for this reason magnetrons are designed to oscillate at the (non-degenerate) π -mode.

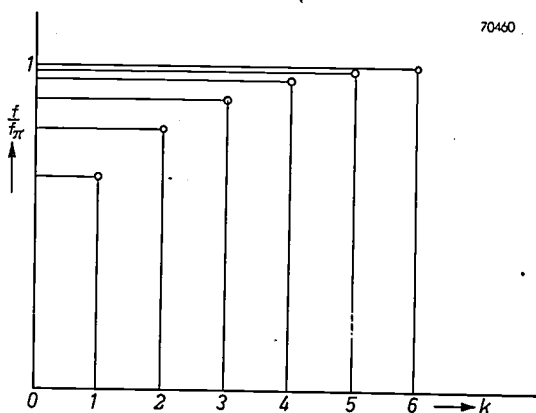


Fig. 15. Natural frequencies of a magnetron with twelve identical cavity resonators. The number k of the mode of oscillation is plotted horizontally and the frequency f vertically, the unit being the frequency f_π of the π -mode ($k = 6$). In the vicinity of the π -mode the frequencies are close together.

Fig. 15 gives an idea of the natural frequencies of a magnetron with twelve cavity resonators; the frequency unit is the frequency of the π -mode. It will be seen that the natural frequencies are very closely spaced in the region of the π -mode, and the risk that the required oscillation will excite the system in such a way as to produce other modes of oscillation is very great. This is detrimental to both the energy transfer and the efficiency, and efforts have therefore been made to find means of increasing the difference between the natural frequencies of the π -mode and neighbouring modes of oscillation.

Improvements in cavity resonator systems

One method of effecting such improvement consists in fitting "straps" between those anode segments which carry the same H.F. voltage, for the π -mode; in other words, all the even segments are strapped with copper wires or rings and likewise the uneven segments. Two methods of strapping are illustrated in fig. 16.

From the aspect of the π -mode, these straps increase the capacitance of the segments with respect to each other. Only the charging current of the capacitance flows in the segments, since the extremities are at the same potential. Let the

combined effective capacitance of all the cavities be denoted by C_t and that of the straps by C_s ; the wavelength is now increased in proportion with $\sqrt{1 + (C_s/C_t)}$, which immediately follows from the well-known expression for the resonant frequency f_0 of an L - C circuit, viz:

$$\omega^2 LC = 1,$$

where $\omega = 2\pi f_0 = 2\pi c/\lambda$ (c = velocity of propagation, λ = wavelength).

For other modes of oscillation, the voltage between adjacent segments is lower and the effective capacitance C_s therefore smaller. (The same applies when extra capacitance is added to a transmission line; the effect of this capacitance is greatest at the voltage antinode and decreases towards the node.) Furthermore, current now also flows in the straps, seeing that the extremities are no longer at the same potential. Effective self-inductance then occurs in parallel with the original inductance and thus reduces the overall inductance. Since both capacitance and inductance are lower than in the case of the π -mode, the wavelength is also smaller.

In this way it is possible to make a fairly clear distinction between the natural frequencies of the π -mode and those of other modes of oscillation, amounting to 5% or more, and it is accordingly possible to prevent oscillation in undesirable modes.

Another method of modifying the spectrum of natural frequencies consists in making the cavities of unequal sizes, the most usual

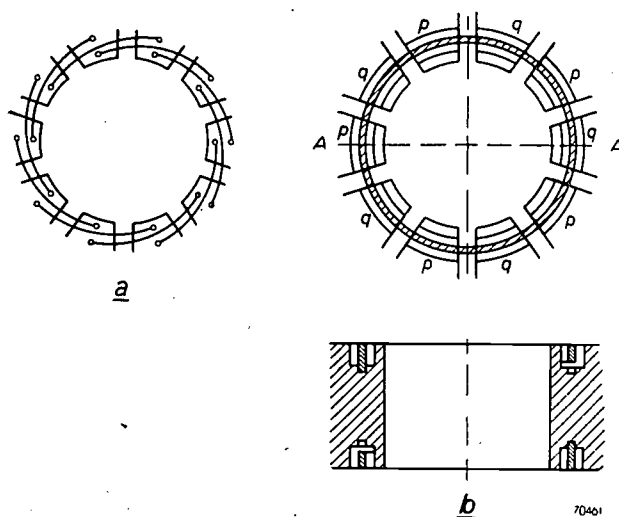


Fig. 16. a) "Echelon strapping". Each anode segment is connected to two others by means of a wire. b) "Ring strapping". The cross section AA is also shown. The ring misses one segment, its height being reduced at that point, and is attached to the next where it is higher. The ring above the anode block connects the segments p , and that below it the segments q .

arrangement being the "rising sun" system depicted in *fig. 17*, which has ten large and ten small cavities.

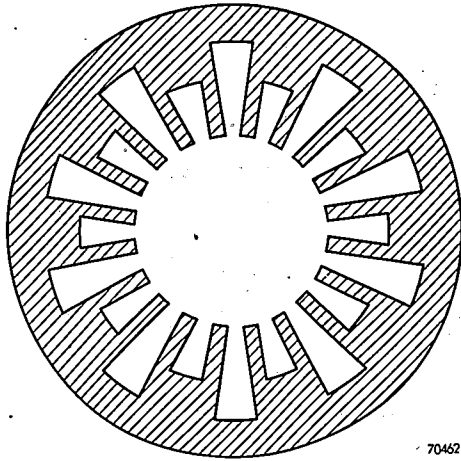


Fig. 17. "Rising sun" anode system with twenty cavities, alternately large and small.

Here, the wavelength of the π -mode lies between two groups, one of which is roughly represented by the natural frequencies of the large cavities only and the other those of the small ones. *Fig. 18* illustrates this in the case of a magnetron having eighteen segments.

It is very difficult — or even impossible — to make strapped magnetrons for wavelengths of less than 1 cm, because the straps would then have to be less than 0.1 mm in thickness. For these millimetric waves, "rising sun" systems are normally used.

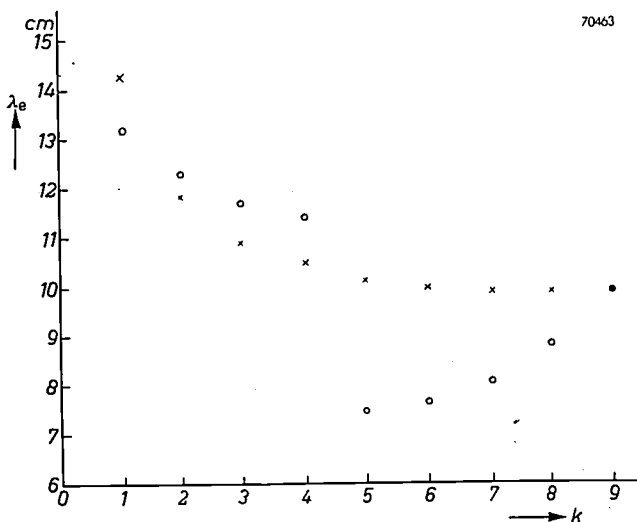


Fig. 18. Natural frequency λ_0 as a function of the number k of the mode of oscillation, for two different anode systems consisting of eighteen cavities, which have the same π -mode wavelength. The crosses refer to an unstrapped magnetron and the circles to a "rising sun" system. (From the previously mentioned article in *Bell Syst. Techn. J.* 25, 229, 1946.)

Output systems

The H.F. power is taken from one of the cavity resonators usually in one of the two following ways.

The first of these employs a loop coupled to one of the cavity resonators and connected to the inner conductor of a coaxial line. Between the inner and outer conductors there is a glass seal which not only provides the output terminal, but also hermetically seals the tube. Often, the coaxial portion is coupled into a waveguide; sometimes, again, the end of the waveguide is soldered direct to the magnetron in the manner shown in *fig. 19*.

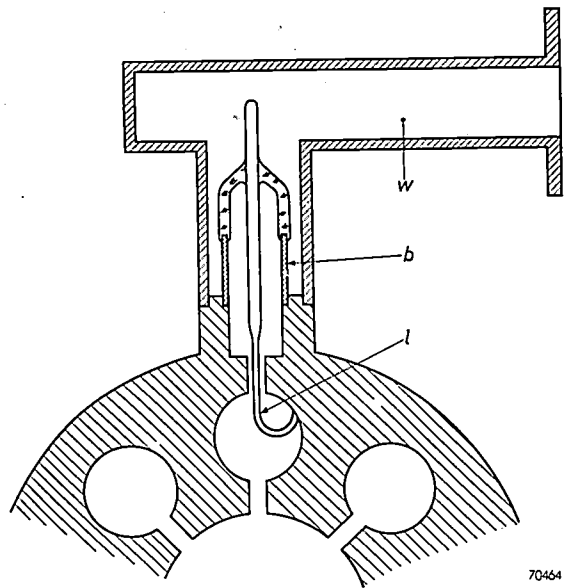


Fig. 19. Coaxial output system. The loop l is connected to an inner conductor which is attached to the outer conductor b by glass. The waveguide w is soldered to the magnetron.

According to the second method a slot is provided in the rear wall of one of the cavities and the power is taken from here direct to a waveguide, which forming an integral part of the magnetron, also contains a "window" which functions as vacuum seal. This method is illustrated in *fig. 20* and it is employed in many instances, e.g. when the power is so high that there would be some risk of flashover to the probe in the waveguide, owing to the high field strength. Or it may be used when the wavelength is so short that it would be difficult to incorporate a loop in the resonators.

The coupling between resonator and waveguide should preferably be such that the greatest possible amount of power is delivered. It should not be too tight, however, for the frequency shift due to small variations in the impedance of the aerial system must remain sufficiently small. Many radar sys-

tems employ a rotating aerial and the rotation always results in some fluctuation in the impedance.

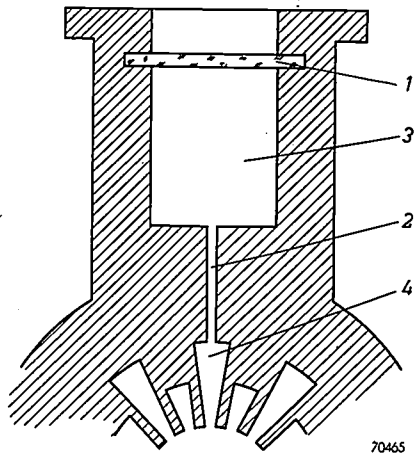


Fig. 20. Cross section of slot coupling (diagrammatic). 1 = vacuum-tight window, 2 = slot, 3 = waveguide, 4 = one of the cavity resonators.

A compromise between the condition for effective transfer of power (tight coupling) on the one hand and that which promotes stability (loose coupling) on the other, is established by suitably dimensioning the coupling gap. In principle it would be possible to provide some matching arrangement outside the magnetron, somewhere in the waveguide, but the great objection to this is that it would produce standing waves in the waveguide, between the magnetron and the matching arrangement; such would tend locally to increase the field strength, which in turn might lead to flashovers. Moreover, it is often required that the magnetron can be replaced quickly, without the necessity of having to make readjustments.

The characteristics of an output system are shown in a so-called Rieke diagram in which the frequency and output power of the magnetron are plotted as functions of the load. Every point in the diagram corresponds to a certain load impedance "seen" from a fixed point in the output line. A rectangular diagram could be employed for this purpose, with the real and imaginary components as co-ordinates, but it is preferable for practical reasons to use the reflection factor and the phase of the reflected wave as variables with polar co-ordinates¹⁰). The reflection factor and phase can, in fact, be measured directly and, moreover, care can be taken in the design of the H.F. lines and aerials that the reflection

factor does not exceed a certain value. The impedance is then kept within a given circle in the Rieke diagram, the importance of which may be explained as follows.

In *fig. 21* the reflection factor is plotted as radius vector and the phase as an angle. Measurements of power and frequency are taken at numerous different values of the impedance (i.e. different sizes of radius vector and angle), and the Rieke diagram is constructed by plotting the curves through the points of constant power or frequency.

It will be seen from *fig. 21* that the power is high in a certain zone, but that the variation in frequency with the impedance is also large. These conditions easily lead to instability. By varying the size of the coupling gap, the zone in question

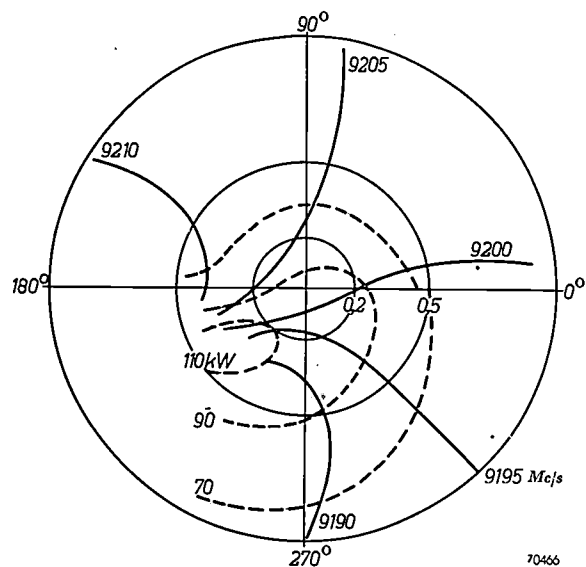


Fig. 21. Rieke diagram of a magnetron for a wavelength of 3 cm. The impedance of the load is plotted in a circular diagram. The reflection factor is represented by the radius vector; the phase of the reflected wave at any point is given by the angle between the radius vector and a fixed direction. In this diagram, lines of constant frequency (in Mc/s) and constant power (in kW) are drawn. The overall frequency variation occurring within the circle for a reflection factor of 0.2 is known as the "pulling figure" (in this case approximately 12 Mc/s).

can be displaced towards the centre, or away from it. The amount of frequency shift is usually indicated by the pulling figure, which is in effect the total frequency variation occurring in the magnetron when the impedance value is moved round the circle representing a reflection factor of 0.2 in the Rieke diagram. Both the pulling figure and the output power are reduced when the width of the coupling gap is decreased. The particular compromise adopted depends on the application.

¹⁰ P. H. Smith, Transmission line calculator, *Electronics* 12, 29-31, 1939.

Constructional details

The cathode

The cathode of a magnetron has to meet some very high standards. In the first place the current density is great; in pulsed magnetrons cathode loads of more than 30 A/cm² are common. Further, at such current densities, in conjunction with anode voltages of 30 kV or more, the cathode must not arc, and it must be possible to dissipate the heat generated by electronic bombardment in the wrong phase (see above). These requirements become so much the more rigid according as the output power is increased and the wavelength reduced.

Oxide-coated cathodes are found to withstand loads of 10 A/cm² surprisingly well, provided these are applied in the form of pulses of about 1 μ sec duration; a cathode life of 1000 hours or more is by no means unusual, but it should be remembered that if the duty cycle is, say, 1 : 1000, a life of 1000 hours means that the cathode has actually emitted only for one hour.

In the case of tubes for still higher cathode loads use is often made of a cathode having metal gauze in the oxide coating, or a metallic powder sintered over it which is coated with another oxide layer. It is actually found that, if this is not done, heavy currents result in such a large potential difference across the cathode that disruptive discharges are likely to occur in the coating¹¹⁾. In consequence, the oxide coating will generally sputter; the gauze or metallic powder more or less short-circuits the coating and thus limits the potential differences.

The L-type of cathode mentioned in an earlier article¹²⁾ is particularly suitable for magnetrons. Not only is the specific emission high, but the resistance of the whole system is, moreover negligible. Arcing is very much less in this type of cathode than in others and the cathode is hardly heated by the pulsatory currents, in contrast with oxide-coated cathodes, in which considerable heat is developed by these currents because of the high resistance of the oxide coating¹³⁾.

In some tubes it is a disadvantage of the L-cathode that it consumes more heater power than an oxide-coated cathode. For the magnetron, however, this is a decided advantage. Once the

tube is working, the cathode remains hot when the heater voltage is reduced to zero, owing to the electrons striking it; it is therefore able to dissipate an amount of power equal to the heater power in the form of back-bombardment heat, which is a certain fraction of the input power. Owing to the higher heater power consumed by the L-cathode, the amount of D.C. input power of a magnetron with this type of cathode can also be higher.

In the case of oxide-coated cathodes, good use is often made of the high secondary emission resulting from the returning electrons. Magnetrons with this type of cathode can be operated at a current that is a multiple of the primary emission. Since the coefficient of secondary emission of the L-type of cathode is very much lower, the use of this cathode means that the dependence on the primary emission is greater.

Fig. 22 depicts the L-cathode as employed in a magnetron suitable for a wavelength of 3 cm. The flanges at the ends are provided in all magnetrons to prevent electrons from straying towards the end spaces of the magnetron itself. Currents arising from such electrons cannot produce any H.F. oscillation and thus result only in detrimental heating of the end plates.

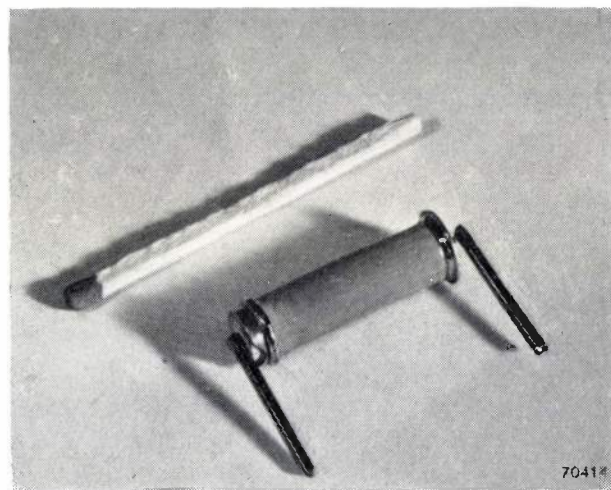


Fig. 22. L-type of cathode for a magnetron for a wavelength of 3 cm.

The anode system

The manufacture of anode systems for magnetrons for wavelengths of 3 cm or less is difficult because of the small dimensions of the system, especially when the magnetron is strapped; the straps, which are not more than a few tenths of a millimetre in thickness, have to be soldered in position very carefully. It will be appreciated that special

¹¹⁾ R. Loosjes, H. J. Vink and C. G. J. Jansen, Thermionic emitters under pulsed operation, Philips Techn. Rev. 13, 337-345, 1952 (No. 12).

¹²⁾ H. J. Lemmens, M. J. Jansen and R. Loosjes, Philips Techn. Rev. II (341-350, 1950).

¹³⁾ E. A. Coomes, The pulsed properties of oxide cathodes, J. appl. Phys. 17, 647-654, 1946.

methods of manufacture have had to be evolved for magnetrons suitable for a wavelength of only 1 cm, the linear dimensions being one third of those of a tube for a wavelength of 3 cm.

There are several ways of making magnetrons with cavity resonators in the form of sectors. The walls, which divide one cavity from the next, can, for example, be soldered to a copper outer ring with the aid of special jigs or, again, the systems may be produced by electrolytic or casting processes, using a suitable matrix.

We shall, however, describe in more detail another method which is particularly suitable when large number of precisely similar anode systems are to be made, even if these be of extremely small dimensions, the method in question being known as hobbing. The anode shown in *fig. 23* has been made in this way. The "hob" consists of a steel "negative" very accurately ground to size. This "hob", mounted in a press, is forced into a solid block of copper, the height of which is roughly equal to the desired length of the anode, and the copper is thus made to flow into the details of the hob, which is subsequently withdrawn.

The grinding of a tool of this kind is a lengthy process; it can be done automatically and with great precision on a machine which each time rotates the hob to the extent of one slot, after

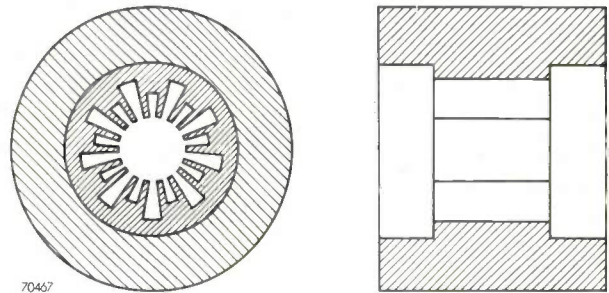


Fig. 23. "Rising sun" anode for a magnetron for a wavelength of 3.2 cm, drawn three-fourths of actual size in cross and longitudinal sections.

which a high-speed grinding disc grinds out the slot to a certain depth. When the hob has completed one revolution the feed is automatically increased and each of the slots is deepened. The leading end of the hob terminates in an obtuse point and the block of copper to be hobbled is recessed to receive

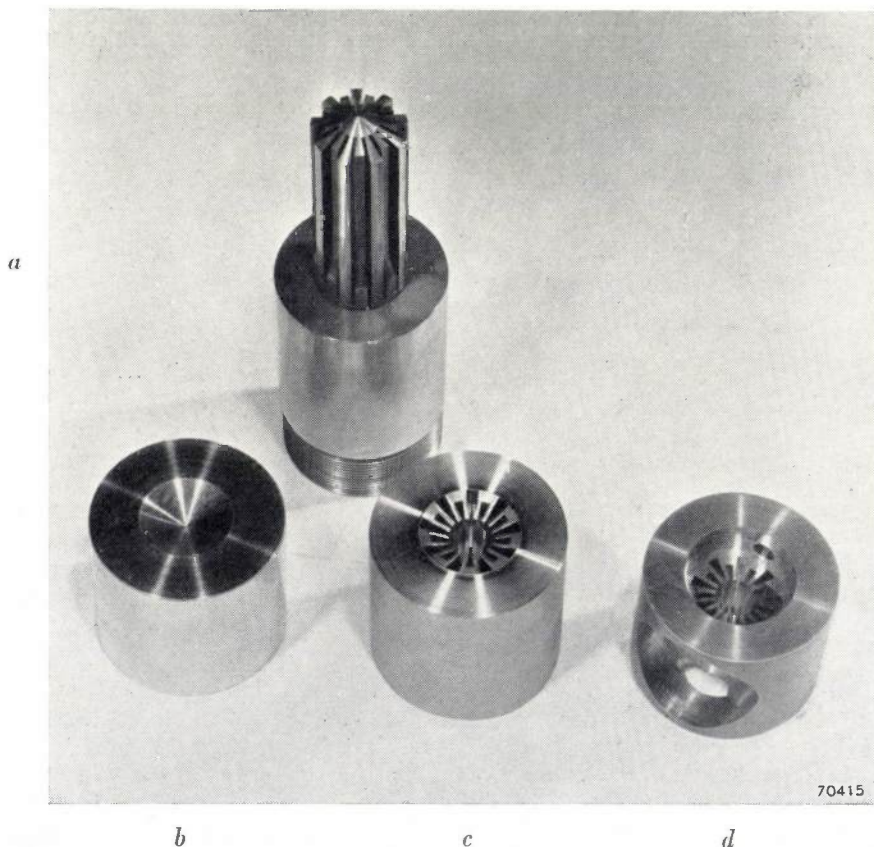


Fig. 24. a) Hob. b) Copper block to be hobbled. c) After hobbing. d) After turning. The waveguide is connected to the aperture in the side in (d).

this. In *fig. 24* the hob is seen at the rear and, in front of it, from left to right, blocks of copper before and after hobbing, and the same block after having been turned down to the required dimensions for the magnetron. In order to protect the cavity walls from damage during the final machining, the hobbled anode is filled with a thermo-plastic substance known as "Lucite", the essential feature of which is that it sets without the inclusion of any air-bubbles. After the turning operation

the "Lucite" is removed by means of a solvent. Some hundreds of anode systems can be made with a single tool of this kind. Incidentally, waveguides and coupling slots can also be produced comparatively easily by the same process.

In conclusion we reproduce in *fig. 25* two photographs of a hobbled magnetron with L-type of cathode, for a wavelength of 3 cm. The performance diagram in *fig. 12* and the Rieke diagram *fig. 21*, relate to this magnetron.

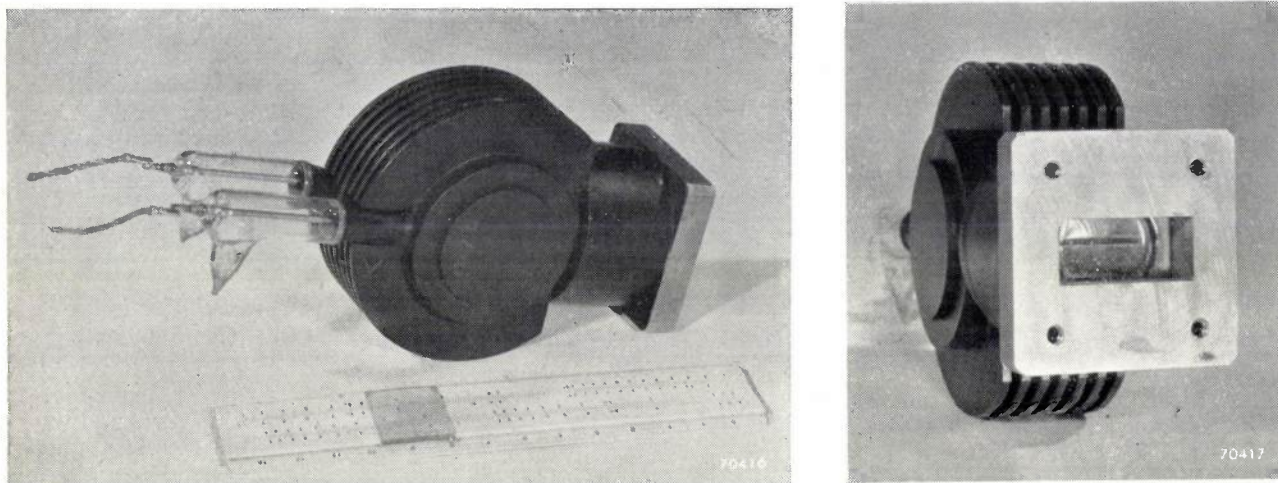


Fig. 25. Magnetron for a wavelength of 3.2 cm, to deliver 1000 kW peak. The right-hand illustration shows the waveguide output connection; the coupling slot is seen through the window. The cathode of this magnetron is depicted in *fig. 22* and the anode in *fig. 24d*.

Summary. This introductory article on magnetrons explains the motion of electrons when influenced by mutually perpendicular electric and magnetic fields. A quantitative explanation of the occurrence of oscillations is given, based on the considerations formulated by Posthumus. A brief account is also given of the useful effects of phase focusing, or formation of electron bunches, and of Hartree's method of iteration as applied to the computation of electronic paths. Different relationships between the direct anode voltage V and the flux density B set a limit in the V - B plane to the zone in which oscillation is possible. Mention is made of the performance chart.

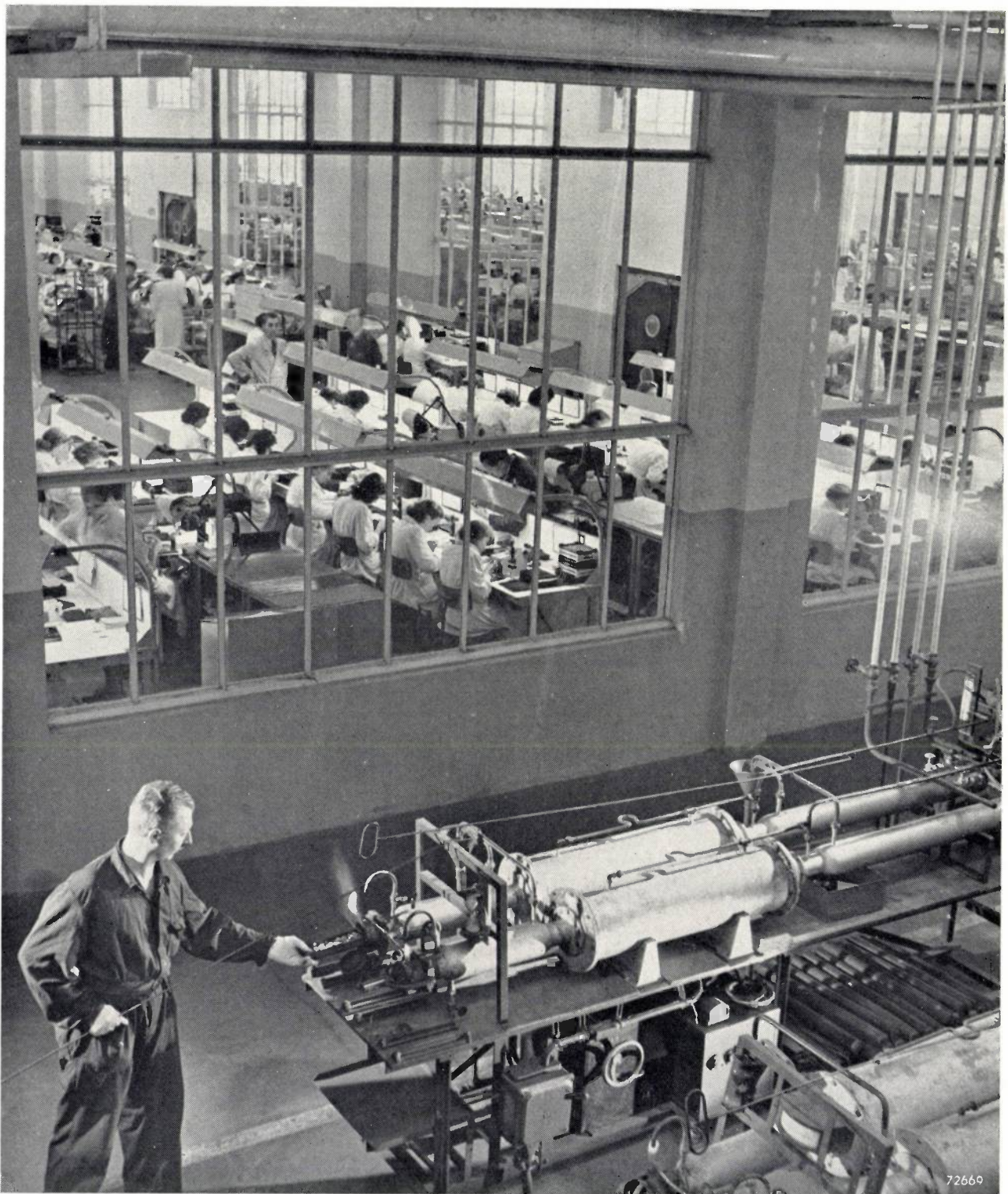
The resonators employed with magnetrons now usually take the form of cavity resonators incorporated in the anode. Details of the spectrum of resonant frequencies resulting from

the coupling between the cavities are discussed, as well as means of increasing the difference between neighbouring resonant frequencies in order to avoid possible unwanted modes of oscillation (strapping, or the use of the "rising sun" system of alternate large and small cavities).

Output systems for magnetrons generally take the form of a coupling-loop or slot in one of the cavities. The characteristics of such systems (inter alia the pulling figure) are expressed in a Rieke diagram, an example of which is given.

Details regarding the design of magnetrons include the stringent requirements to which the cathode must conform (these being adequately met by the L-type of cathode), and the method of manufacture of the anode block with cavity resonators by the hobbing process.

PART OF THE PHILIPS VALVE FACTORY AT SITTARD (HOLLAND)



72669

Photograph Walter Nürnberg

STRAIN AGEING IN IRON AND STEEL

by J. D. FAST.

620.193.918.4:669.11:669.14

There is no need to stress the importance of fundamental research into the influence of impurities in technical metals such as steel. Impurities are responsible, i.e., for the ageing phenomena arising in iron and steel after cold working (rolling, drawing). Experiments carried out with alloys of well-defined composition show once more the important part played particularly by nitrogen as an impurity.

Ageing

Steel, and particularly so-called mild steel, shows ageing phenomena following upon mechanical deformation (rolling, drawing). Such phenomena manifest themselves mainly in a gradual increase in the hardness of the material after deformation. This gradual change in the hardness of iron and steel takes place also after rapid cooling (quenching) from a high temperature. In the former case one speaks of strain ageing, in the latter case of quench ageing. An article on quench ageing appeared in this journal a short time ago¹⁾, when it was explained that this phenomenon is due to the precipitation of particles of iron carbide or iron nitride in the metal, such arising from the presence in the metal of quantities of carbon and nitrogen in the dissolved state prior to the quenching; pure iron does not show any quench ageing. The reader is reminded here of the remarkable phenomenon that a small quantity of manganese appeared to delay quench ageing of nitrogen-containing iron considerably.

Strain ageing, too, appears to be caused by particular constituents in the metal. In *fig. 1* curves are given showing the change in hardness taking place in course of time after quenching and after deformation of mild steel (according to experiments by Davenport and Bain²⁾). The broken lines represent the change in hardness after quenching from 720 °C to room temperature and subsequent storage at temperatures of 40 °C and 100 °C respectively. It appears that the hardness first increases and then decreases. The aforementioned precipitation of iron carbide and iron nitride in the form of small particles causes the increase in hardness, the subsequent decrease being due to coagulation (coarsening) of the precipitated particles.

As the curves show, the higher the temperature the more rapidly both these changes take place.

The fully-drawn curves show the change in hardness after the material has been subjected to a deformation of 10% at room temperature (cold working). In this case, too, there is at first an increase in hardness during storage, the increase being more rapid the higher the temperature, but there is virtually no decrease in hardness as time goes on. Thus in this respect there is a decided difference between strain ageing and quench ageing.

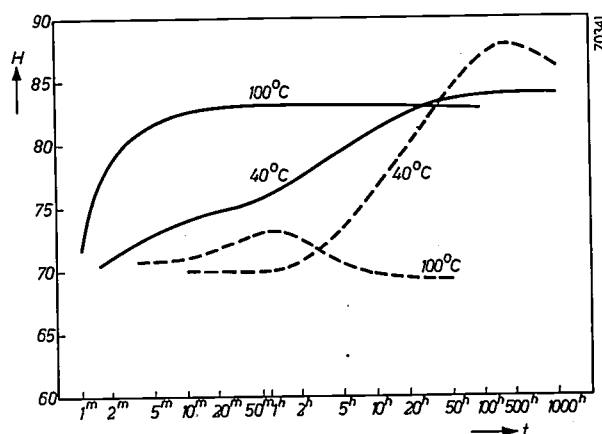


Fig. 1. Quench ageing and strain ageing of mild steel, according to Davenport and Bain²⁾. Vickers hardness H as a function of the time t at 40 °C and 100 °C, after quenching from 720 °C (broken-line curves) and after 10% cold deformation (fully drawn curves). The hardness of the quenched material first increases (precipitation) and then decreases (coagulation). The hardness of the deformed material increases, but at the temperatures given does not subsequently decrease. Note the difference in maximum hardness at the two temperature in the case of the quenched material. For the deformed material the final value of the hardness is practically independent of the ageing temperature.

Another effect of strain ageing is to be seen in the stress-strain curve of technical iron. This curve represents the relation between the load (stress) to which the rod-shaped material is subjected and the deformation (strain) resulting therefrom. In *fig. 2a* a stress-strain curve is given for mild steel.

¹⁾ J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips Techn. Rev. 13, 165-171, 1951 (No. 6). This article will further be referred to as I.

²⁾ E. S. Davenport and C. E. Bain, Trans. Amer. Soc. Met. 23, 1047, 1935.

Under small loads there is an almost linear relation between the elongation (deformation) and the load, but when a load is reached corresponding to the point *A* plastic deformation of the material begins to take place, the deformation then increasing

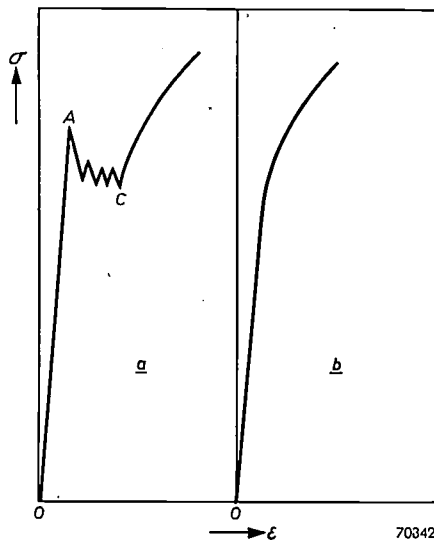


Fig. 2. Stress-strain curves of mild steel. Stress σ plotted as a function of the deformation ϵ . *a*) The steel shows an upper yield point at *A*. Material first strained as far as *C*, shows under renewed loading a smooth stress-strain curve (*b*).

very suddenly and rather irregularly, while at the same time the load decreases somewhat as a result of the inertia of the tensile testing machine. Finally, beyond *C*, any further plastic straining of the material requires again an increasing load. The stresses corresponding to the points *A* and *B* are called respectively the upper and lower yield points of the steel.

When the load is removed as soon as the point *C* in fig. 2*a* is reached and the stress-strain curve is recorded anew, the smooth curve of fig. 2*b* is obtained. The yield points have disappeared. Strain ageing then manifests itself in a gradual return of the sharp yield points when, following upon the first treatment (deformation up to the point *C*), the material is left to itself, and the higher the temperature at which the material is kept the more quickly this return takes place.

Neither of these two effects of strain ageing — gradual hardening after deformation and gradual return of the two yield points — occurs in pure iron (which as a matter of fact shows no yield points at all). It is therefore to be concluded that strain ageing as well as quench ageing is related to impurities which are always present in technical iron. But whereas the cause of quench ageing has been found to lie in the formation of precipitated

particles, according to Cottrell³) strain ageing is to be ascribed to the behaviour of the carbon and nitrogen atoms remaining in the dissolved state.

To check this assumption, hardness tests have been carried out with pure iron and with iron to which known quantities of one or more impurities had been added by a method described earlier in this journal⁴); as impurities, carbon (0.04%) and nitrogen (0.02%) were used, with and without manganese (0.5%). Before proceeding to discuss the experiments it is well to consider briefly the explanation suggested by Cottrell *et al* for strain ageing, since the results of our experiments seem to make it necessary to reconsider some details of that theory.

Plastic deformation of a metal

The plastic deformation of a metal consists in parts of crystals shearing over each other along certain crystallographic planes (slip planes) and in certain crystallographic directions (slip directions) under the influence of external forces. Fig. 3 shows, diagrammatically, how this might happen. The broken line represents a slip plane perpendicular to the plane of the drawing, the slip direction being along the broken line. Under the influence of the shear stress indicated by arrows, shearing takes place in the manner represented in figs 3*a* to 3*c*. In fig. 3*c* the shear has progressed over the distance of one atom, the atoms meanwhile having passed through positions (fig. 3*b*) in which they have a greater energy than before or after the slip. This energy can be calculated from the crystal characteristics

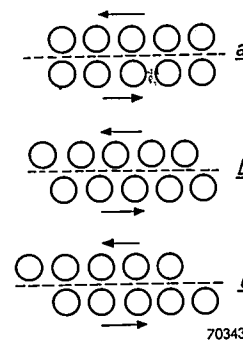


Fig. 3. How two layers of atoms in an imaginary simple cubic metal lattice glide continuously one over the other. To reach the new position of equilibrium *c* from the position *a* the atoms have to pass through the configuration *b*, which has a much greater energy.

³) A. H. Cottrell, Report of a conference on strength of solids, held at Bristol, 1948, pp. 30-38; Progress in Metal Physics 1, 77-126, 1949.

⁴) J. D. Fast, Apparatus for preparation of metals with an exactly known content of impurities, Philips Techn. Rev. 11, 241-244, 1949.

and from that it can be deduced what shear stress would be needed to bring about a displacement of all atoms simultaneously, as represented in fig. 3. Such a calculation gives stresses more than a thousand times greater than those found experi-

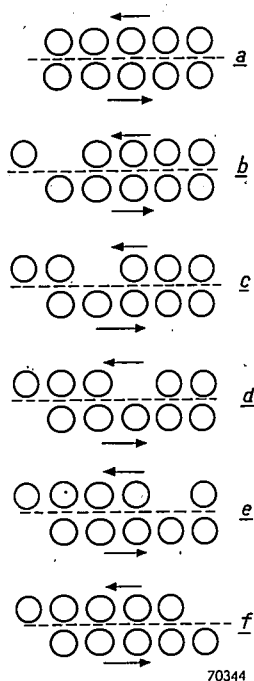


Fig. 4. Simplified representation of the discontinuous gliding of one layer of atoms over another in a simple cubic lattice. The arrows in this and the previous diagram denote the direction of the shear stress. The broken line represents the slip plane.

mentally, so that apparently the atoms are not all displaced simultaneously. The shear might be imagined as taking place in steps, the atoms being displaced in the slip direction one by one, in which case a much smaller shear stress is required. Represented in the most primitive form, the situation then changes from that in fig. 3 to that in fig. 4.

The representation in fig. 4 is rather too simple, since only those atoms are considered which lie immediately above and below the slip plane. When taking also the other atoms into account we get the picture of fig. 5, where the shear has progressed half-way through the crystal (thus the situation as given in fig. 4c or fig. 4d). The deviation from the ideal structure as shown in the middle of fig. 5 (the outlined part) is called a dislocation. It is to be borne in mind that here only a cross section through the crystal is shown and that the dislocation is assumed to extend perpendicularly to the plane of drawing on either side. The dislocation as it exists in the middle of the diagram, while proceeding from left to right, can also be imagined as being formed locally owing to an extra, vertical,

layer of atoms perpendicular to the paper (indicated by the short arrow) being forced into the lower half of the crystal. As a result the atoms lie too close together in the lower half of the dislocation and too far apart in the upper half, the lower half being under a compressive stress while the upper half is under a tensional stress. In addition, shear stresses arise round the dislocation.

To the left of the dislocation a slip has already taken place over the distance of one atom, while to the right of it there is as yet no change.

In this picture plastic deformation is nothing but the movement of dislocations along the slip planes in certain crystallographic directions. Under the influence of a shear stress this movement takes place at a high speed, being promoted by the thermal motion of the atoms. When a dislocation has been displaced from one side of the crystal to the other, a shear has taken place over the distance of one atom. Viewed from above a slip plane, the shearing process takes place as diagrammatically represented in fig. 6. The (hatched) dislocation must either extend over the entire width of the slip plane from one side of the crystal to the other or else form a closed loop.

It is commonly assumed that there are always dislocations already present in the metal crystals.

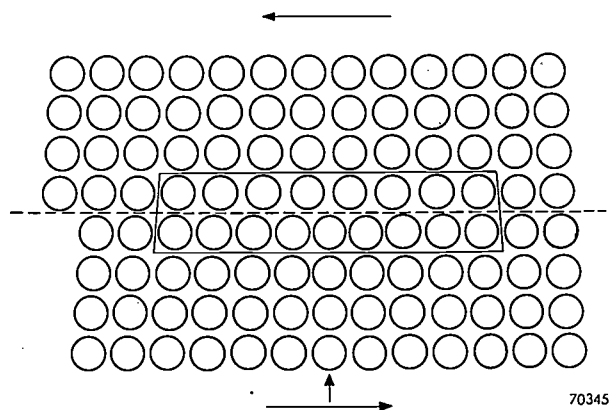


Fig. 5. Schematic representation (in one atomic plane) of a dislocation in a simple cubic metal lattice. The horizontal arrows denote the direction of the external shear stress, the broken line representing the slip plane. In the outlined area the influence of the dislocation is most manifest. The dislocation can also be imagined as being caused by the forced entry of an extra row of atoms in the lower half of the crystal (short arrow). The dislocation is a practically linear lattice disturbance perpendicular to the plane of drawing; here only the cross section with one atomic plane is to be seen.

How it is to be imagined that in the formation of the crystals from a melt or from a gas phase these dislocations become incorporated in the solid substance, and how they can be formed at the walls of a crystal during the process of deformation, are

questions which we shall not enter into here. Anyhow it must be assumed there is a great increase in the number of dislocations during the plastic deformation, and this accounts for the hardening of the metal as a result of the deformation. Each dislocation being surrounded by a stress field, the dislocations exercise attractive or repelling forces upon each other, and as they increase in density the movement of each individual dislocation is rendered more difficult by these forces. Thus the metal becomes less malleable and its hardness increases. This hardening is clearly expressed in the stress-strain curves of fig. 2.

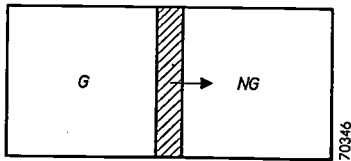


Fig. 6. Displacement of a dislocation from left to right along a slip plane, "seen" in a direction perpendicular to that of the three preceding diagrams, viz. in the direction of the short arrow in fig. 5. *G* is the part already sheared ("the part where the dislocation has already glided over"), *NG* is the part not yet sheared.

According to the dislocation theory the increase in the hardness of a metal after quenching is due to the movement of the dislocations being hampered by the precipitated particles, this probably being caused mainly by the elastic deformation of the lattice surrounding the grains of the precipitate.

Theoretical considerations on strain ageing

According to Cottrell *et al*, strain ageing results from the interaction between the dislocations and the carbon and nitrogen atoms interstitially dissolved in the iron. As already explained in article I, each dissolved carbon or nitrogen atom increases the volume of the elementary cell of the iron lattice in which it is contained. Now in the foregoing we have seen that on one side of a dislocation the iron lattice is stretched and on the other side compressed. Consequently the interstitial places are no longer of equal energy: the dissolved atoms escape from the compressed areas and tend to seek the expanded areas. In the state of equilibrium at room temperature only few atoms are available for escaping from the compressed regions, since the solubility of carbon or nitrogen in iron at room temperature is very small indeed, only a few times 10^{-7} and 10^{-5} wt% respectively. Therefore the interaction between the dislocations and the dissolved atoms will mainly be such that in course of time in the expanded part of each dislocation a concentration

of C or N atoms much greater than that in the undisturbed lattice will be established. This results in a decrease of the tensile and compressive stresses surrounding the dislocation. The shear stresses, too, may be reduced by displacements of the dissolved atoms, since each C or N atom causes an asymmetric deformation of the iron lattice (see I). This may all be briefly expressed by saying that the elastic energy of the dislocation (or rather, of the stress field around the dislocation) is reduced.

In the picture of fig. 5 the cloud of C or N atoms formed in and around a dislocation might consist mainly of one central row of atoms, i.e. at most one C or N atom per atomic plane parallel to the plane of drawing. Around that "cloud" the concentration decreases exponentially. Such strings of C or N atoms may perhaps be regarded as an intermediate stage between the dissolved and precipitated states of carbon (nitrogen). It might be assumed that these one-dimensional pre-precipitates act as nuclei in the precipitation proper.

As a consequence of its reduced elastic energy the dislocation becomes, as it were, anchored, and thus the hardness of the material is increased. In this way strain ageing may be considered as a process of diffusion of the interstitially dissolved carbon or nitrogen atoms towards the dislocations.

In a technical steel that has not recently been deformed a large portion of the dislocations will always be anchored. The displacement of such a dislocation will increase its elastic energy, unless it is able to take along with it its cloud of interstitial atoms. When at a relatively low temperature a comparatively rapidly increasing load is applied to the material, the clouds will not have time to diffuse and as the load increases the dislocations will finally be "torn away" from their clouds, after which a smaller stress is required for the further movement of the dislocations. To put it in other words, a greater stress is required to initiate the plastic flow than to keep the material flowing. Thus the theory gives a natural explanation for the occurrence of the two yield points in technical iron (see fig. 2a).

If after the first overall yielding the iron is not further deformed and the stress-strain curve is again recorded after not too long a time, the dissolved atoms will not yet have had sufficient opportunity to diffuse towards the dislocations again, so that the latter will then still be able to move easily. This explains why no yield points are found in the new stress-strain curve. If, however, one waits so long, after the recording of the first curve, that the interstitial atoms have had sufficient time to diffuse towards the dislocations and thus to anchor them, then it is to be expected that an

additional stress will again be needed to start plastic flow, in other words, that the two yield points will have returned. As already stated, this is exactly what is observed. The higher the temperature, the greater will be the rate of diffusion and the sooner will be the return of the yield points. This, too, is confirmed experimentally, as we have already seen.

The theory is quantitatively confirmed by the fact that the temperature coefficient for the rate of this strain ageing in carbon-containing iron appears to be about equal to the temperature coefficient for the diffusion constant of carbon in iron (corresponding to an activation energy of about 80 000 joules per mole, see I).

Strain-ageing experiments with specially prepared alloys

It was considered well worth while to investigate whether the above-mentioned views held by Cottrell and others, which are mainly based upon experiments with technical iron, also hold in the case of pure iron to which known quantities of one or more impurities have been added. In the course of the experiments carried out with these materials⁵⁾ it appeared that carbon and nitrogen do not have exactly the same effects. Even after forty hours' heating at 50 °C carbon did not give any strain ageing, whereas nitrogen gave the maximum effect already after two hours' heating at that temperature. The differences are shown graphically in *figs 7 and 8*. With carbon as the only impurity the iron has to be heated to 200 °C (heating time two hours) before the maximum ageing effect is obtained.

The procedure for these experiments was analogous to that followed for the quench-ageing experiments described in article I. First the hardness of the material was measured immediately after the deformation and then, for the ageing phenomenon to take place within a reasonable time, the material was kept at 50 °C for two hours, after that at 100 °C for two hours, at 200 °C for two hours and so on up to 700 °C, taking measurements of the hardness after each period of heating. The advantage of this procedure is that the last stage of the ageing, which at low temperatures sometimes takes a very long time, now takes place at higher temperatures, thus requiring less time.

Figs 7 and 8 show that after the heat treatment at 300 °C the hardness of the material decreases again. This would appear to contradict our statement that, as opposed to what takes place in quench ageing, in the case of strain ageing there is no loss in hardness after long storage. This, however, is only a seeming contradiction. At temperatures above 300 °C the processes of recovery and recrystallization come into play and it is this that causes the hardness to be reduced. The descending part of the curves in *figs 7 and 8* is therefore due to processes other than those dealt with here.

⁵⁾ J. D. Fast, *Revue de Métallurgie* 47, 779-786, 1950.

The experiments described indicate that Cottrell's theory, according to which strain ageing of technical iron is caused by a regrouping around the dislocations of carbon or nitrogen atoms which are already dissolved in the iron, has to be modified slightly.

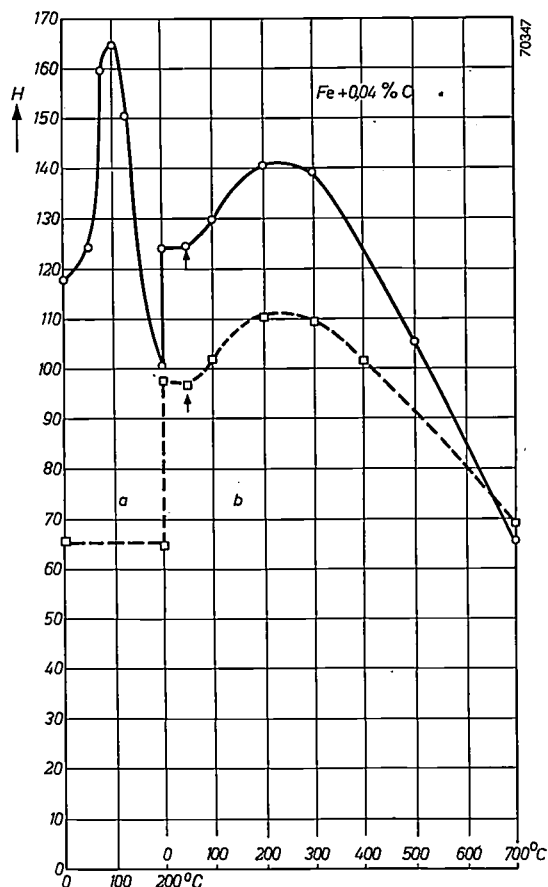


Fig. 7. The Vickers hardness H of iron containing 0.04% carbon as the only impurity, after different thermal and mechanical treatments. The fully-drawn curve gives the values of the hardness after the following treatments: quenching from 720 °C followed by two hours' heating successively at 50, 75, 100, 125 and 200 °C, 10% deformation (elongation), two hours' heating at 50, 100, 200, 300, 500 and 700 °C. The broken curve gives the hardness after deformation alone and heating at the various temperatures for two hours. In neither case is there any perceptible strain ageing (see the small arrows) during the two hours' heating at 50 °C. The part *a* of this diagram relates to the change in hardness after quenching, part *b* to the change in hardness after cold deformation.

Cottrell's theory can certainly account for the fact that at low temperature carbon does not immediately cause strain ageing. It is known that carbon is much less soluble in iron than nitrogen: the solubility of nitrogen in iron is about $2 \times 10^{-5}\%$ at 20 °C, whereas in the case of carbon this value is only reached at 150 °C. Apparently, therefore, below about 100 °C there are not sufficient carbon atoms in solution for the dislocations to be anchored. The fact, however, that at a somewhat higher temperature, say 100 °C, carbon already gives rise

to a relatively rapid strain ageing, indicates that atoms must then be available for anchoring dislocations, whilst the slightly increased solubility at that temperature is certainly not sufficient to account for the effects observed. It is, therefore, to be concluded that the carbon atoms required for anchoring the dislocations are supplied for the greater part by the carbide particles. Apparently these carbon atoms pass into solution and then through diffusion move towards the dislocations. This would imply that from the point of view of energy it is at least just as advantageous for the C atoms to be in a dislocation cloud as in a carbide crystal.

According to this modified theory the ageing in carbon-containing iron will take place at the same rate as that in nitrogen-containing iron at 50 °C, only at that temperature where the product of the diffusion coefficient and the solubility of carbon is equal to that for nitrogen at 50 °C. It is thereby

presumed that the distances between the dislocations and the grains of the precipitate in the carbon-containing iron are about equal to those in the nitrogen-containing iron, and that the transition of a C atom from the precipitate to the solution requires about the same activation energy as the corresponding transition of an N atom.

According to the most recent figures⁶⁾ the solubility q and the diffusion coefficient D of carbon and of nitrogen in alpha iron are given by the formulae:

$$q_C = 3,1 e^{-10\,000/RT} \text{ wt } \%,$$

$$q_N = 5,0 e^{-7\,100/RT} \text{ wt } \%,$$

$$D_C = 0,02 e^{-20\,100/RT} \text{ cm}^2 \text{ sec}^{-1},$$

$$D_N = 0,003 e^{-18\,200/RT} \text{ cm}^2 \text{ sec}^{-1},$$

where T stands for the absolute temperature and R is the gas constant per mole. According to these formulae the product $D \cdot q$ for nitrogen in alpha iron at 50 °C is

$$D_N q_N (50 \text{ }^\circ\text{C}) = 1,4 \cdot 10^{-19} \text{ cm}^2 \text{ sec}^{-1}.$$

The product $D_C q_C$ reaches this value at a temperature of about 100 °C. Actually, however, the ageing in carbon-containing iron at 100 °C still takes place at a slower rate than that in nitrogen-containing iron at 50 °C (figs 7 and 8). This could be explained by assuming that more nitrogen was dissolved in the iron than corresponded to the state of equilibrium, or, in other words, that we had to do with supersaturated solutions of nitrogen, in contrast to what was the case with the carbon-containing specimens. Under such conditions the nitrogen could anchor the dislocations much sooner than the carbon, which first has to pass into solution from the carbide particles. Another possible explanation is that the transition of a nitrogen atom from the precipitate to the solution takes place more readily than the corresponding transition of a carbon atom.

In the foregoing calculation q_N relates to the solubility of nitrogen in alpha iron which is in equilibrium with the stable nitride phase Fe_4N . In the material that was used for the ageing experiments the nitride was actually in the form of the metastable phase Fe_{16}N_2 (cf. I). The solubility of nitrogen in alpha iron which is in equilibrium with the latter phase is greater and according to Dijkstra⁶⁾ is given by

$$q'_N = 32 e^{-7800/RT}.$$

Applying this expression in the above arguments, the temperature of about 100 °C referred to becomes about 105 °C, so that the result changes but little.

⁶⁾ For the solubility figures see L. J. Dijkstra, Tr. Am. Inst. Min. and Met. Engrs 185, 252-260, 1949, and for the diffusion coefficients see C. A. Wert, Phys. Rev. 79, 601-605, 1950 and J. Appl. Phys. 21, 1196-1197, 1950.

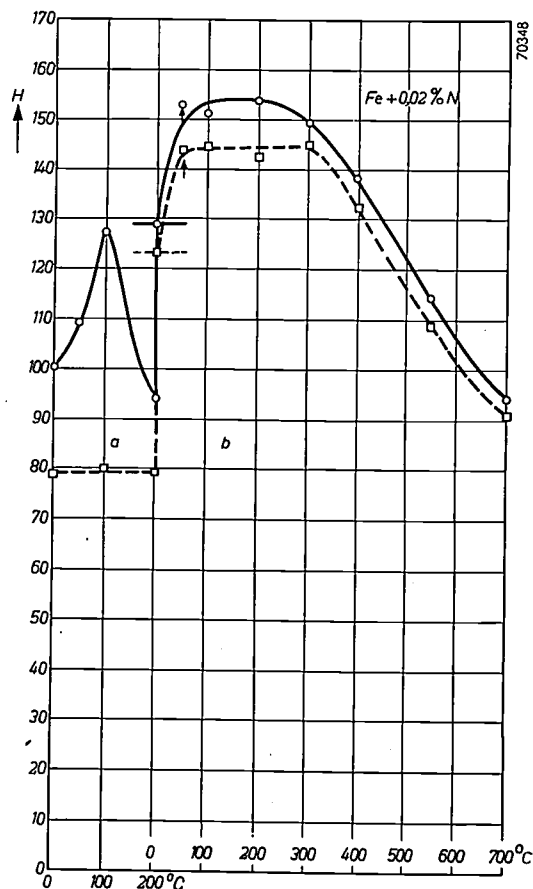


Fig. 8. Vickers hardness H of iron with 0.02% nitrogen as the only impurity, after the thermal and mechanical treatments defined in the subscript of fig. 7. The parts denoted by a and b have the same significance as in fig. 7. Strain ageing takes place very much quicker with nitrogen than with carbon as impurity; the maximum hardness is already reached after two hours' heating at 50 °C (small arrows). The decrease in hardness at higher temperatures in these two diagrams is due to recovery and recrystallization.

The experiments lead to the conclusion that as an impurity in mild steel nitrogen is to be regarded as being very much more dangerous than carbon. This is also borne out by the fact that by adding aluminium a technical steel can be produced which shows very little strain ageing (e.g. the German Izett steel). Aluminium binds only the nitrogen contained in the steel, and not the carbon⁷⁾.

The effect of manganese upon strain ageing

As described in I, the addition of 0.5% manganese to a nitrogen-containing iron delays its quench ageing considerably. This was accounted for by the fact that the dissolved nitrogen atoms become strongly bound to the manganese atoms in the iron lattice, so that they cannot readily precipitate⁸⁾. But, remarkably enough, the results of our experiments on strain ageing have shown that manganese has no perceptible effect upon the strain ageing of nitrogen-containing iron. From the experiments on quench ageing it would have been expected that manganese prevents the dissolved nitrogen from diffusing to the dislocations and that consequently nitrogen would hardly be able to cause strain ageing in an iron containing manganese.

It is difficult to reconcile this phenomenon with the foregoing considerations. A direction in which the solution might be found lies in a possible interaction between the manganese atoms and the dislocations, such that after the deformation a considerable portion of the manganese atoms is present in the dislocations. It is also possible that our explanation for the effect of manganese upon quench ageing is not quite correct and that the delaying action of manganese in the formation of nitride nuclei is to be accounted for in some other way than by supposing that the nitrogen atoms are unable to release themselves readily from the manganese atoms.

Blue brittleness

Some of our experiments concerned with the problem of blue brittleness are even more difficult to fit into the dislocation picture. By the

blue brittleness of technical iron (mild steel) is to be understood the fact that at 200 °C—300 °C this material shows a smaller elongation and reduction of area and a greater tensile strength than at lower or higher temperatures (see *fig. 9*). When the

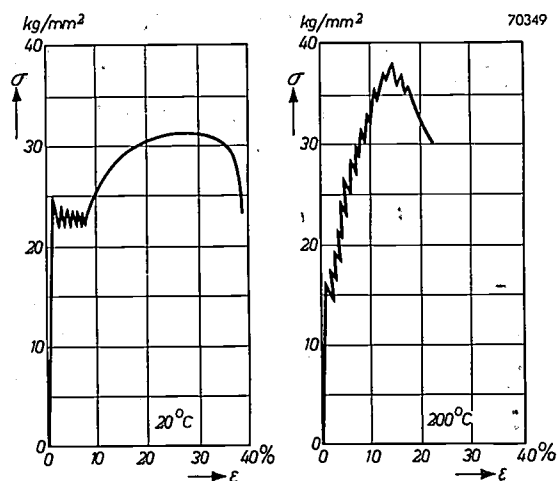


Fig. 9. Stress-strain curves of a non-killed mild steel, taken at 20 °C and at 200 °C respectively. The material shows less elongation and greater tensile strength at 200 °C than at 20 °C. The name of blue brittleness given to this phenomenon is related to the heat tinting occurring at 200 °C. The elongation has been measured as the change in length in percents of a test bar originally 5 cm long.

test is carried out at a faster rate this state of relative brittleness is shifted to a higher temperature. An example of such a test is the impact test⁹⁾, where a notched bar of given dimensions is broken with a hammer and the labour required to do that is measured. The impact value of technical iron as a function of temperature shows a minimum at temperatures between about 400 °C and 500 °C. Our experiments (described in the article⁹⁾) show that there is no such minimum at all in the curve for the impact value when carbon is the only impurity in the iron. With nitrogen as impurity, on the other hand, there is indeed a minimum (see *fig. 10*).

With the aid of the dislocation theory¹⁰⁾ an attempt has been made to explain the blue brittleness and the minimum in the curve for the impact value in the following way. When the deformation takes place very quickly or very slowly the dislocations experience little hindrance from the clouds surrounding them. When the rate of deformation is very fast the clouds are unable to keep pace with

⁷⁾ According to investigations by S. T. Epstein, H. J. Cutler and J. W. Frame, *J. Metals*, June 1950, pp 830-834, strain ageing can also be suppressed by adding vanadium instead of aluminium to the steel. Vanadium combines with carbon as well as with nitrogen, but the object aimed at was reached with much less vanadium than would have been required to bind all the carbon in addition to the nitrogen.

⁸⁾ J. D. Fast and L. J. Dijkstra, *Internal friction in iron and steel*, *Philips Techn. Rev.* 13, 172-179, 1951 (No. 6).

⁹⁾ See J. D. Fast, *Investigations into the impact strength of iron and steel*, *Philips Techn. Rev.* 11, 303-310, 1950.

¹⁰⁾ See also F. R. N. Nabarro, *Report of a conference on strength of solids*, held at Bristol 1948, pp 38-45, and G. Masing, *Arch. Eisenhüttenwesen* 21, 315-325, 1950.

the dislocations and thus no longer have any effect after the latter have broken away from them, whilst in the other case, when the rate of deformation is very slow, the clouds have little effect because they can quite easily keep pace with the movement of the dislocations. There are, however, intermediate rates of deformation where the clouds are still able to move with the dislocations, but only with difficulty. In this range the dislocations are impeded in their movement, and this finds expression in reduced ductility of the material. The higher the temperature, the greater is the diffusion rate of the C and N atoms and the easier the clouds are able to keep pace with the dislocations. As the

temperature rises the region of brittleness is therefore shifted to higher rates of deformation, or, what amounts to the same thing, the quicker the deformation takes place the higher is the temperature at which brittleness occurs. At first sight this explanation would seem to be acceptable. It is not clear, however, why nitrogen should cause the material to become brittle and carbon not.

Although, therefore, the dislocation theory gives many useful indications about the influence of certain impurities in technical iron, in its present form it does not yet explain the details of the behaviour of the impurities.

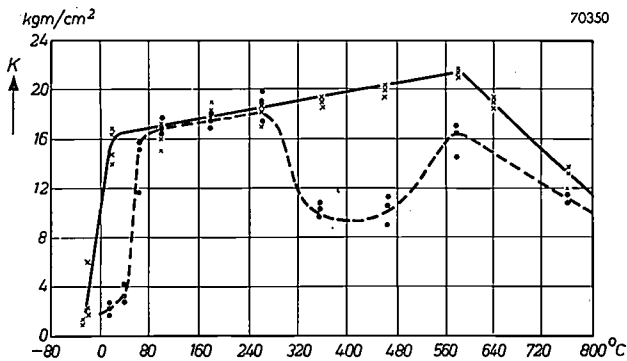


Fig. 10. When carbon is the only impurity in the iron (the fully drawn curve relates to iron with 0.02% C) no minimum occurs in the curve representing the impact value K as a function of temperature. On the other hand, if both carbon and nitrogen are contained in the iron (the broken-line curve relates to iron with 0.06% C and 0.018% N) there is a minimum in the impact value curve between 400 $^{\circ}\text{C}$ and 500 $^{\circ}\text{C}$.

Summary. By strain ageing of iron and steel is understood a gradual increase in hardness after cold working and the reappearance of the two yield points in the stress-strain curve after its disappearance through pre-treatment. The ageing phenomena are to be ascribed to impurities such as carbon and nitrogen contained in the metal. With the aid of the theory of dislocations attempts can be made to explain these phenomena (Cottrell), but the difficulty is that in the state of equilibrium the amounts of dissolved carbon and nitrogen are too small to cause the effects observed. It is therefore to be concluded that carbon and nitrogen pass into solution from precipitated particles of carbide and nitride respectively and then flow towards the dislocations. Experiments with carefully prepared alloys containing only carbon or nitrogen as impurity show that even after forty hours' heating at 50 $^{\circ}\text{C}$ carbon still does not cause any strain ageing, whereas already after two hours' heating at that temperature nitrogen has the maximum effect. Apparently nitrogen has a much greater effect upon the properties of iron than carbon, as is also apparent from experiments concerning the blue brittleness and the impact value of steel. As far as strain ageing is concerned, manganese does not affect the behaviour of nitrogen, a very remarkable phenomenon, considering that quench ageing is indeed greatly influenced by manganese.

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.

1987: W. de Groot: A graphical method for solving problems of low-level photometry (Appl. sci. Res. B 2, 131-148, 1951, No. 2).

A certain level of subjective brightness may be characterised by the minimum power per unit area and solid angle, E_m , with which the corresponding brightness impression can be evoked monochromatically, the corresponding wavelength being λ_m . If the visibility function $V(\lambda_m, \lambda)$ for that level of subjective brightness is known, it is possible to find the luminance $L_E = K \int E(\lambda) V_p(\lambda) d\lambda$ of a radiation $E(\lambda)$ for which $\int E(\lambda) V(\lambda_m, \lambda) d\lambda = E_m$, giving the same brightness impression as E_m . This

is achieved by a graphical method which can be incorporated in a slide rule for low-level photometry. Luminance scales for different relative energy distributions are easily obtained and, after choosing one of them as a standard of subjective brightness, other luminances may be corrected according to the standard scale chosen. Luminances for monochromatic radiations of wavelength λ may be obtained, in the same way. When plotting these versus λ for a number of values of E_m , Bouma's diagram is obtained, in which the data for composite radiations may also be plotted. An analysis is given of Weaver's data regarding $V(\lambda_m, \lambda)$, and the

various functions discussed in the preceding sections are calculated using these data. Finally the function $E_m(\lambda_m)$, as obtained from these calculations, is checked with other data from literature.

1988: J. J. Went and H. P. Wijn: The magnetisation process in ferrites (Phys. Rev. **82**, 269-270, 1951, No. 2).

Below the gyromagnetic resonance frequency the initial permeability of ferrites is, in general, independent of frequency. From this it has already been concluded that the magnetisation at very low field strengths must be governed entirely by rotational processes. Measurements at higher field strengths, (up to $\mu_0 H = 10^{-4}$ Wb/m² ($H = 1$ Oe)), and as a function of frequency up to 1800 kc/s, have shown that a second mechanism exists, which (a) is frequency-dependent at frequencies below the gyromagnetic resonance frequency, (b) takes place at a lower field strength as the density of the ferrite is greater, (c) is irreversible. The additional process is ascribed to irreversible displacements of the Bloch walls.

1989: F. A. Kröger and P. Zalm: The origin of multiple bands in luminophors activated by divalent manganese (J. Electrochem. Soc. **98**, 177-182, 1951, No. 5).

The two principal theories which have been advanced for the explanation of the appearance of multiple bands with manganese-activated luminophors, particularly the silicates and germanates, are the "cluster theory" and the "co-ordination theory". The former accounts quite naturally for the dependence of the red/green ratio on the Mn concentration, whereas the co-ordination theory requires both the effect of co-ordination and concentration quenching (or possibly some other factor) for a self-consistent picture. On the other hand, the effect of Be in promoting the red emission is quite naturally explained by the co-ordination theory, whereas the cluster theory requires an additional postulate concerning the effect of Be in promoting clusters. Taking these and other considerations into account, the cluster theory seems to present the simpler and more satisfactory picture. In the cluster theory, the variation of luminescence observed is attributed to the interaction between manganese ions belonging to clusters and it seems that this influence must affect mainly the lifetime of the various excited states. It seems probable that the fluorescence bands are due to

electronic transitions from various excited states to one single ground state and that several different bands may originate in one centre. The variation of fluorescence with temperature can be explained on the basis of this picture.

1990*: P. Cornelius: Kurze Zusammenfassung der Elektrizitätslehre. Eine Einführung des rationalisierten Giorgischen Maßsystems (Vienna, Springer Verlag, 1951; VI + 89 pp., 11 figs, 7 tables). (A short survey of the theory of electricity, an introduction of the rationalized Giorgi system of electromagnetic units; in German).

For the contents of his booklet the reader is referred to abstract No. 1803*. The German edition is extended with a chapter on the difference between field strength and induction in the electric and the magnetic field and further contains an "addendum", in which the author deals briefly with possible objections of physicists regarding his way of exposition. An English edition is in preparation.

1991: J. L. Snoek: Conférence sur les propriétés magnétiques des ferrites (J. de Physique Radium **12**, 288-238, 1951, No. 3). (Conférence on magnetic properties of ferrites; in French).

Survey of magnetic properties of ferrites, in which the following points are discussed: Fe₃O₄, ferrites XY₂O₄, ferrite technology, normal and inverted spinels, Néel's theory of antiferromagnetism, variation of saturation magnetisation with temperature, magnetic resonance, magnetic crystal energy, magnetostriction, dispersion and absorption at high frequencies. To this paper a report of E. W. Gorter has been added, which deals with an experimental illustration of the influence of the angle Fe-O-Fe on the indirect exchange energy.

1992: A. Bierman and W. Hondius Boldingh: The relation between tension and exposure times in radiography (Acta Radiologica **35**, 22-26, 1951, No. 1).

Experiments, carried out under widely varying conditions, have proved that the density of a radiograph remains practically constant when varying kV and mAs in such a way that the value of the product (kV)⁵ × mAs is kept constant. Application of this rule does not give intolerable deviations in the whole field of diagnostics.

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

WAYS TO COMFORTABLE LIGHTING

by L. C. KALFF.

628.97

By its basic problem and by the way the solution of this problem is approached this article is characteristic of the personality of the writer, much more so than are most other articles published in this Review. Ir. Kalff, who is well known in many countries for his achievements in architectural and exhibition lighting etc., has propounded at several international meetings in the past year the views expressed in this article. The editors feel that these views, balancing on (or trespassing beyond) the borderline between science and art, may be of interest to the layman as well as to the expert in the lighting field.

The problem of comfortable lighting

Up to 1910-1920 it was usual to have in our houses rather dark walls and ceilings. Artificial lighting was, in comparison to what we have now, scarce and for that reason could only be used as local lighting. A lamp over the table, a reading lamp, wall brackets near the fireplace could spread only small circles of light.

Gradually the output of electric lamps increased. Electric current prices went down and light became cheaper. From 1915-1925 the first general lighting systems appeared in offices and factories. It is true that the levels used then would now be considered much too low (4 to 6 lm/sq.ft). In those days we find the tendency to use white ceilings and light colours for walls and furniture. This trend had been advertised not only by decorators but it was also warmly supported by lighting men who stressed the point of efficiency.

Gradually this vogue to have light colours in interiors became more and more general in Europe and in America. After a while, however, it was found that high level installations (e.g. 50 lm/sq.ft) did not give satisfactory results in these light coloured interiors. This fact was put forth in 1937 by Ward Harrison of Cleveland in his remarkable article¹⁾: "What is wrong with our fifty-footcandle installa-

tions?" The question was answered by the observation that in many high level installations glare or a feeling of discomfort is caused by the light sources, or that the attention is distracted by an unsuitable distribution of brightnesses in the visual field. A new conception, called "Brightness Engineering", was to direct the activity of the lighting engineer. This conception gained even more importance by the advent of the fluorescent lamps with their high efficiency, causing a very rapid development in lighting technique and lighting levels since 1945.

The aim of the conception of Brightness Engineering was mainly in the negative: discomfort had to be avoided. The following theories aim at making possible a positive approach to the problem. It is not enough to avoid discomfort but we should endeavour by careful planning to create comfort. It will be made clear in this article that this setting of the task should lead to closer co-operation of the lighting expert with the creative artist, architect and decorator, as we will have to go back to darker and coloured walls and ceilings in order to obtain similar comfort and atmosphere to that which we had before, with the lower level installations.

Such a co-operation will perhaps not easily be established. Up to now co-operation between the two groups has been an almost completely one-way

¹⁾ W. Harrison, Ill. Eng. 32, 208, 1937.

traffic. Lighting men were trying to induce the artist into including light in his decorations, so as to obtain a certain lighting level, avoiding discomfort from glare, shadows or reflections. The artist has been reluctant to respond to these efforts, for various reasons. Obviously he does not like to complicate still more his work which is already tremendously complicated with modern techniques and constructions. Further, the methods and instruments offered to him by the lighting expert, useful though they may be in the hands of the latter, are mostly very unhandy tools in the hands of the creative artist. But above all, the artist rightly must have assessed the tools given to him by the lighting expert as inadequate for his aims. In fact, for the architect — and really for everybody — the calculations of brightness and brightness ratios and also of glare are of a very limited value, since these calculations can only be made after conception of an interior. In the activity of the creative artist they can furnish a check but no guiding indications in what direction to think.

There seems to be only one way to improve this situation: only when the artist feels that he can handle light in much the same way as he can handle colours, shapes and materials, will the lighting problem cease to be a nuisance to him. He will feel able to master the problem of seeing-comfort as soon as this will have become part of his natural activities. In order to favour this development the lighting expert should take a completely new attitude towards the study of lighting. He should try to understand the problems of the artist and consider the way in which a lighting system together with the decorator's shaping of a room can provide seeing-comfort.

This is the basic trend of the new approach towards these problems proposed in this article and on several occasions before²⁾.

In presenting this approach to readers of this Review, we are well aware that our reasoning is by no means strictly scientific. When entering the realm of the artist one must not be surprised that in many cases belief will be substituted for knowledge, and, in fact, many things suggested here are difficult to prove in the sense that we can prove a mathematical or physical law.

A scheme of lines, colours and brightnesses

If we want to observe a visual task in comfort, we may analyze our requirements into two groups.

²⁾ L. C. Kalf, Conditions for comfortable seeing, Lecture given at I.C.I. Conference, Stockholm 1951, published also in *International Lighting Review*, Vol. 1951/52, No. 2, 19-26. Comfortable lighting, Lecture given at Meeting of I.E.S. of Great Britain, Eastbourne 1952.

Firstly, we want to see the essential characteristics of the task, the shape, the colour, and the dimensions, and for that reason we sometimes want shadows, highlights or a high level of lighting, or a special colour of light. Secondly, however, we also want to see the task in a concentrated way, i.e. we do not want to be distracted by influences around it.

Now every picture we see, looking at a visual task, consists of three elements: lines, colours and brightnesses, and these three elements are apt to render vision either easy or uncomfortable. Let us consider these three elements separately.

1) Lines

The lines in the picture we see, may be arranged in a rhythmic and quiet way or they may be chaotic. Let us think for instance of looking at a wood of high beech trees, with parallel trunks, and looking at it when we are lying on our back, seeing the criss-cross design of the branches overhead.

The lines in a picture can concentrate our attention on the task, either by pointing towards it or by being arranged concentrically around it. If there are many lines that have not been arranged so and have no relation whatever to the seeing task, they will not be co-operative in concentrating our attention on what we want to observe.

2) Colours

The colours may be divided into two groups: the warm ones, from yellow through orange to red, purple and brown, and the cool ones, from yellowish green to green, blue, violet, grey and black.

The warm colours seem to be the most attractive to us. They also seem to many of us to be nearer than the cold ones. For that reason we should like to concentrate the warm colours on the visual task and its immediate surroundings, whilst we think the cool colours are more adequate for the outer parts or what may be called the background in the visual field.

In practice we unconsciously very often arrange colours in this way. An actress or a singer in a theatre will prefer to appear in a bright and warm-coloured dress against a cool back-ground. One can hardly imagine an artist on the platform in a grey or blue dress against a bright red or orange screen. If we want to decorate the centre of a room by putting flowers on the table, we do not nearly obtain the same vivid impression from blue flowers in a vase on red tablecloth as we do from red, orange or yellow flowers on a green cloth.

3) *Brightness*

It is generally accepted that the area having the greatest brightness in a picture catches the eye. We can find many examples for this. We put an artist in the limelight, we lay the dining table with a white cloth, the bride is dressed in white. Napoleon chose to sit on a white horse during his battles, etc.

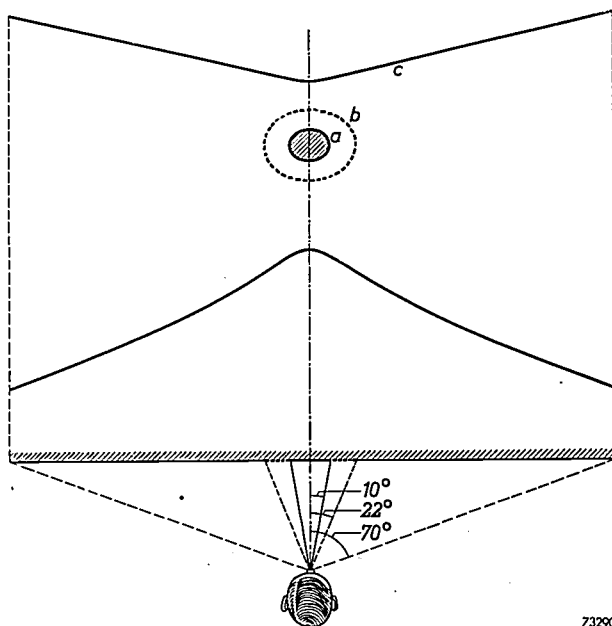
Thus it is only logical that, if we want to keep concentrated on the visual task, this task should be the brightest object in the visual field, its immediate surroundings should be not so bright and the background even less so.

If we now compose a scheme for these three elements in the visual field, we get a picture in which lines, colours and brightnesses co-operate to keep the attention of the observer concentrated on the task.

Of course, when envisaging the practical application of such a scheme, the question arises as to which part of the visual field should be considered as belonging to the "task", which to the "immediate surroundings" and which to the "background". No exact delineation of these parts can be given, but reasonable limits are found by considering the movement of the head and the eyes of a person looking at some object. It is found that an elliptic cone with horizontal apex angle of about $2 \times 10^\circ$ and a vertical apex angle of about $2 \times 8^\circ$ includes all the objects that are examined with moving eyes, without moving the head (*fig. 1, a*). This is the approximate

angle subtended by an open book of normal size seen from a normal reading distance. This part of the visual field may be regarded as the object of the desirable concentration, i.e. as the task. If an object of rather large size, say, a life-sized portrait, is presented to the average observer, it is found that in order to get a first impression of the picture he involuntarily chooses a viewing distance of about 2.5 times the width or about 3.2 times the height of the picture. Thus, when directing his eyes to one edge of the picture the opposite edge will be situated at an angle of 22° or 17.5° respectively from the viewing direction. The cone determined in this way, of horizontal apex angle $2 \times 22^\circ$ and vertical apex angle $2 \times 17.5^\circ$, may be considered to include the task with its "immediate surroundings" (*fig. 1, b*). That this cone of vision is not arbitrary but has a real psychological significance is corroborated by the fact that this solid angle approximately corresponds to that part of the retina within which the perception of red and green as colours is possible. Outside of this part of the retina there are only rods, which see differences in brightness but cannot distinguish colours nor perceive small details. This outer part of the field of vision then, represents what in the above was called the "background" and which may be assumed to be of relatively small importance in comfortable seeing (*fig. 1, c*). We know that vivid brightness contrasts and quick movements can easily be registered from these angles, but that we have to turn our head in order to see what really is the cause of the distraction.

Having settled in a rough way the question of the definition of task, surroundings and background, we are confronted with another, more difficult question: whilst the arrangement of lines and colours is mainly a qualitative matter, the brightnesses of necessity involve a consideration of quantity. Our general scheme only recommends the establishment of suitable brightness differences for concentrating the attention on the centres of interest. Obviously there should be minimum brightness contrasts serving the purpose; a white room with white furniture and a very diffuse lighting is a completely impossible surrounding for comfort. On the other hand, the brightness contrast must not exceed a certain limit, in order to avoid discomfort by glare, etc. Which, then, are desirable and which maximum permissible brightness ratios? A general rule in Brightness Engineering states that the brightness ratio of simultaneously seen objects should not exceed 1:3. Though this rule may be of some help, we are pretty sure that in some cases larger ratios may be applied without impairing comfort: a good impression seems



73290

Fig. 1. Parts of the field of vision. *a* area of the visual "task", *b* immediate surroundings of the task, *c* background of the visual field. In the latter area colours and small details cannot be perceived. The curiously curved edges of the visual field are due to the cut-off by forehead, nose and cheeks.

largely to depend on the abruptness of brightness variation. In order to get an idea of what is comfortable in this respect, we have constructed spectacles with different sets of glasses, in which light absorbent materials form differently shaped optical wedges, circular and parallel. Experimental work with these spectacles is still in progress.

Practical application of the scheme

The fundamental scheme for lines, colours and brightnesses proposed in the above, of course, will have to be adapted to circumstances. Consider, first, a single person working at a desk. In general, nobody does keep his eyes fixed on the same object for a very long time. Even a person who is writing or reading looks up now and then and likes to fix his eyes on other things, which are a little more distant. These things then become another seeing-task which should present a similar arrangement of lines, colours and brightnesses, in order to be interesting enough. That means a decoration scheme in which there are two centres of interest (*fig. 2*).

For rooms in which people sit, looking in different directions, like dining rooms, board rooms, etc. the scheme has to be altered into horizontal zones of different brightness and colour. For instance in a restaurant the zones of interest in which the seeing-

tasks can be found are the dining table and the faces of the other guests. Over and under these zones the brightness and the warmth of the colours should diminish. For such kinds of rooms it will be advisable, *inter alia*, to keep the ceiling rather dark, either by keeping the light away from it or by giving it a dark colour. Indirect lighting in a restaurant destroys all atmosphere and takes away the attention from the essential things in the room.

Very similar considerations apply to a lounge or lobby: the primary task in this case is found in a horizontal zone, viz. the faces of the other guests; on the other hand, when relaxing or interrupting a conversation, a secondary area of interest should be provided, which obviously should cover again a horizontal zone. An admirable example of a room decorated and lighted in conformity with these views is found in the lobby illustrated in *fig. 3*. The second zone of interest is formed by the grand mural, with a dark panelling underneath and a dark ceiling.

In a church we are inclined to look up. There we may speak of vertical zones, the central one of which has to be brightest and warmest in colour, whilst the zones further away from the centre have to become gradually darker and cooler. *Fig. 4* is a photograph of a church which very distinctly exhibits this scheme of colour and brightness distribution.



Fig. 2. Decoration scheme of a study or office. The person working at the desk cannot keep her eyes fixed on the seeing-task on the desk for a very long time. She will look up now and then to rest her eyes, for which purpose a second centre of interest is provided by a brightly coloured and well lit picture on the wall. The second centre of interest may also be a view through a window, a bunch of flowers or similar things.

Many more examples might be given in order to illustrate our point. In fact, in the decorating practice in Europe of the last five years remarkable instances can be indicated in which the lighting and decorating problems have been tackled with as a background the line-colour-brightness schemes described in this article. Because of our

colour-brightness theories were the principles on which not only the lighting system was based, but which also to a great extent directed the choice of colours and brightnesses of walls, floor and ceiling. It is obvious that in a store the goods to be sold must be the principal items of interest. To this end, local lighting of show-cases and shelves along the



Fig. 3. Lobby of the Automobile Club of Argentine, Buenos Aires. The faces of the guests will constitute a horizontal zone of interest, whilst for relaxation a second horizontal zone of interest is found in the great mural. The concentration on this zone of interest is enhanced by the dark panelling underneath and the dark ceiling.

limited space we can mention only two other cases. *Fig. 5* is a photo of a show-window in which the arrangement of lines, all pointing towards one of the centres of interest, greatly contributes to the attention value of this window. Of course, this is not a typical example of the problem of "seeing-comfort" but it is very typical indeed of one side of this problem as analyzed above, viz, the easy "concentration" of the mind of the onlooker. An example encompassing a greater array of structural problems is illustrated in *fig. 6*. This photograph shows part of a big department store (Vroom and Dreesmann) at Rotterdam. In this project the light-

walls provides the necessary high brightness; this lighting is supported by the concentrated beams of reflector lamps mounted in the ceiling in the louvres which shield the fluorescent lamps giving a general lighting. The rest of the ceiling is completely dark, being painted in night blue and receding into almost complete invisibility behind a grid of light coloured metal strips. This type of ceiling has proved very satisfactory and in recent years has grown rather popular: many decorators in Europe have adopted the "Rotterdam ceiling" when building new or refashioning old department stores. The importance of not distracting the attention by a bright ceiling



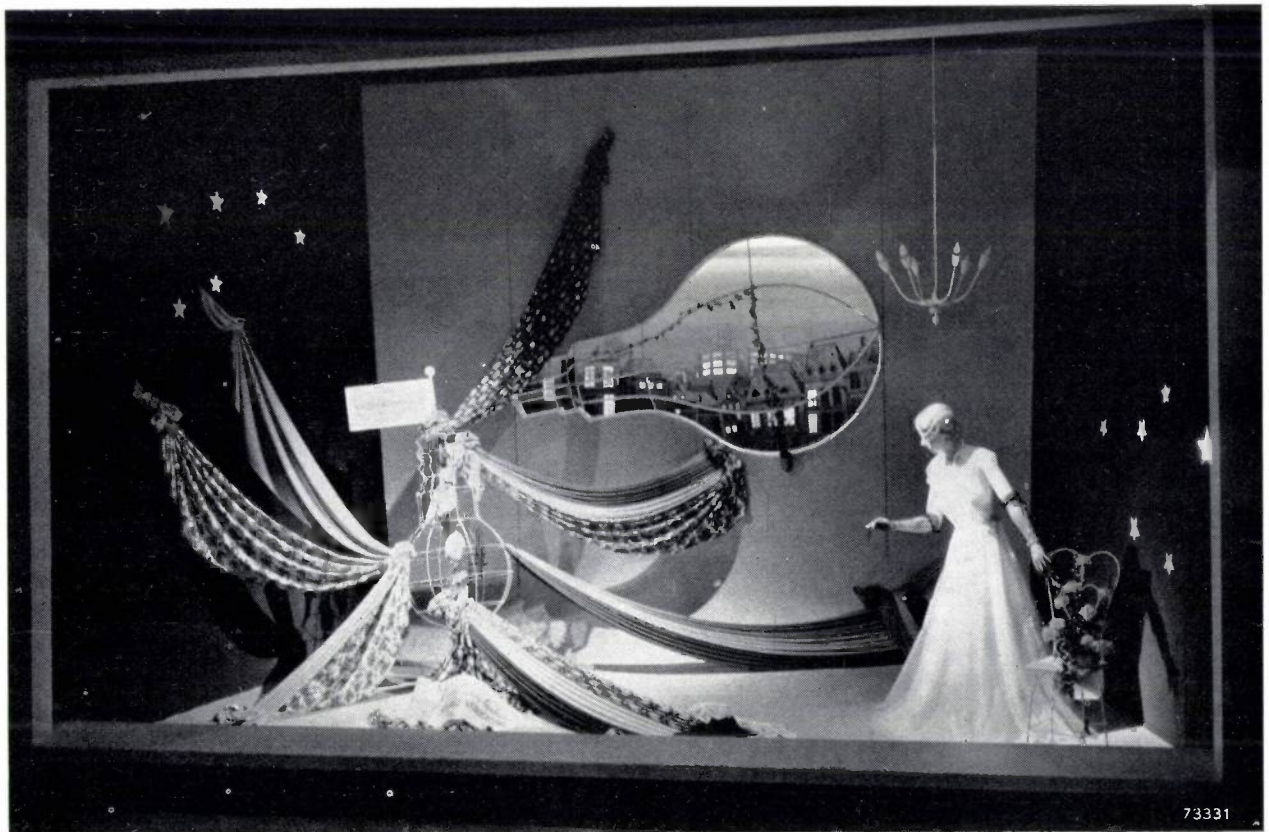
73330

is clearly demonstrated by *fig. 7*, showing a department store decorated on similar lines as the former, but having a ceiling that is painted white and rather brightly illuminated. There can be no doubt that in this case the attention of the visitor is directed to a lesser degree onto the goods in the store.

Gathering of evidence for desirable schemes

It would have been quite natural to end the article at this stage. A theory has been put forth, practical applications of the theory are described and it is shown that satisfactory results are obtained. However, it must be admitted that the theory is not founded on facts but rather on intuition. Such a peculiarity should not prevent the artist from bringing the theory to life in his creations. It never did. But on the other hand, lighting experts generally are of a more scientific mind and will feel a dire need of some evidence of the soundness of the proposed scheme.

Fig. 4. Interior of church at Venray (Netherlands), showing vertical zones, with highest brightness and warmest colours (white, red and gold) in the central zone; in the outer zones dark stone columns with shadows prevail.



73331

Fig. 5. Show-window, containing several centres of interest. By the lines pointing to the chief centre and by a distribution of colours and brightness according to the scheme exposed in this article a high attention value of this window is obtained.

For their purpose the evidence found in the results of applications of the theory will not suffice, as these results quite naturally also depend on many other factors. More direct evidence, obtained in experiments eliminating most of the incidental and complicated influences occurring in practice, may be

persons to whom seeing and constantly analyzing what they see has been the most important part of their activities, and who have more or less consciously played on human reactions to visual impressions: the painters. Now, it is indeed a very interesting experience to go through a museum



Fig. 6. Part of the department store of Vroom and Dreesmann at Rotterdam. The merchandise is made the centre of interest throughout by local lighting of a high level. The ceiling is painted dark blue. In each quadrangle of the ceiling either a lighting unit, a loudspeaker or a clock can be mounted. The general lighting is effected by means of groups of two 40 W fluorescent lamps with two 68 W incandescent reflector lamps employed as ballasts, each group being inserted in two adjacent quadrangles of the ceiling (in this part of the store double groups are used). As the ceiling itself, lying above the bright metal strips, is almost completely invisible, the architects obtained here an easy method of concealing and at the same time keeping accessible the wiring and plumbing of the store, as well as the ducts for air conditioning. This "Rotterdam ceiling" has since been adopted by a score of department stores in Europe.

desirable. In the following a few efforts to produce such evidence will be described. Strange though it may seem to bring forth evidence for a theory only after its extensive application, we feel justified in this procedure, as we do not wish to blur the role of intuition in our approach to the lighting problem.

Since time immemorial there has been a group of

of pictures and to analyze in how far the different works of art are in accord with the principles exposed in this article. We have found that most of the pictures which have been composed as free works of art, inviting quiet contemplation and which have got no other aim (as e.g. giving a mental shock), exhibit more or less clearly our fundamental

schemes of colour, line and brightness. An example is shown in *fig. 8*³⁾.

This is still a kind of indirect evidence. Experiments with the aim of obtaining direct evidence for our views were set up in the "Demonstration



Fig. 7. Department store of Vroom and Dreesmann at Haarlem. The decorating and lighting system is similar to that of Rotterdam (*fig. 6*), but the ceiling is given a brightly painted surface with a high lighting level. The concentration of attention on the merchandise in this case is in our opinion markedly less.

Laboratory" at Eindhoven, inaugurated in May 1951. A few of the facilities of this Laboratory are shown in the following illustrations.

Fig. 9 is a booth in which the different decoration schemes are made visible in a niche. They can be illuminated with different intensities and are in the shape of little metal doors, so that the different schemes can be observed one after the other by simply turning over the panels like the pages of a book. In succession schemes are shown that are and others that in one way or another are not in keeping with our theories, and the preference of the observer is noted.

Fig. 10 shows an entire room in which the colour and lighting scheme can be changed. The walls

consist of turning panels, so that they can be made white, a dull green or in elmwood. The tops of the tables can revolve, and so take the colours of the elmwood or the green walls. Several paintings are hung in this room: the front sides in a colour and line composition obeying the rules of our scheme, the backs in an opposite arrangement. By changing the lighting systems and the colours and reflections of walls and paintings, different aspects and atmospheres can be offered to the spectators in this room, so that a preference for certain combinations of colours and brightnesses can be traced.

We cannot dwell on other possibilities of experimentation provided in the Demonstration Laboratory. Suffice it to say that many hundreds of persons have gone through the experiments, contributing by their reactions to our knowledge of the subject. On the whole these reactions give us reason to believe that we are working in the right direction.

It must be well understood that what we have done here is not an attempt to give recipes for artistic effects. We think that our lines-colours-brightness schemes may be useful because they respond to the way man is thinking and subconsciously reacting to the lines, colours and brightnesses he takes in through the eyes. The schemes are simple and understandable for both lighting expert and artist, but they are so vague as to leave enough freedom to the creative mind. Their chief importance may well be that they will convince the artist of the fact that "light is part of the picture" and so should be part of his activity, whilst on the other hand they will show the lighting man that a lighting system or a lighting level are not things to be planned apart and without intense co-operation with the artist.

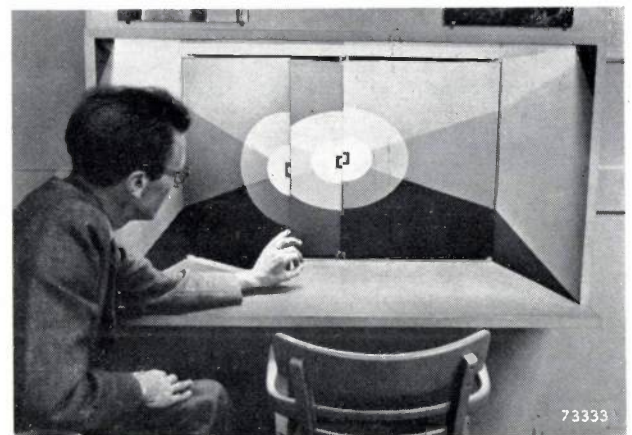


Fig. 9. Niche in the Philips Demonstration Laboratory at Eindhoven, where a number of different lines-colours-brightness schemes can be presented to a spectator.

³⁾ Other examples are to be found in the article in *International Lighting Review* quoted in footnote ²⁾.



Fig. 8. The doge Leonardo Loredan, by Giovanni Bellini (about 1430-1516). This famous portrait exhibits very clearly a scheme of lines, colours and brightnesses, concentrating the attention on the warm-coloured face with the strong contrasts of the dark eyes. Lines of the collar and the hat with its ribbons are arranged concentrically around it, the lines of the shoulders and the row of buttons point towards it and the blue-green (originally pure blue) background is cool and receding, whilst its brightness rapidly diminishes near the frame.

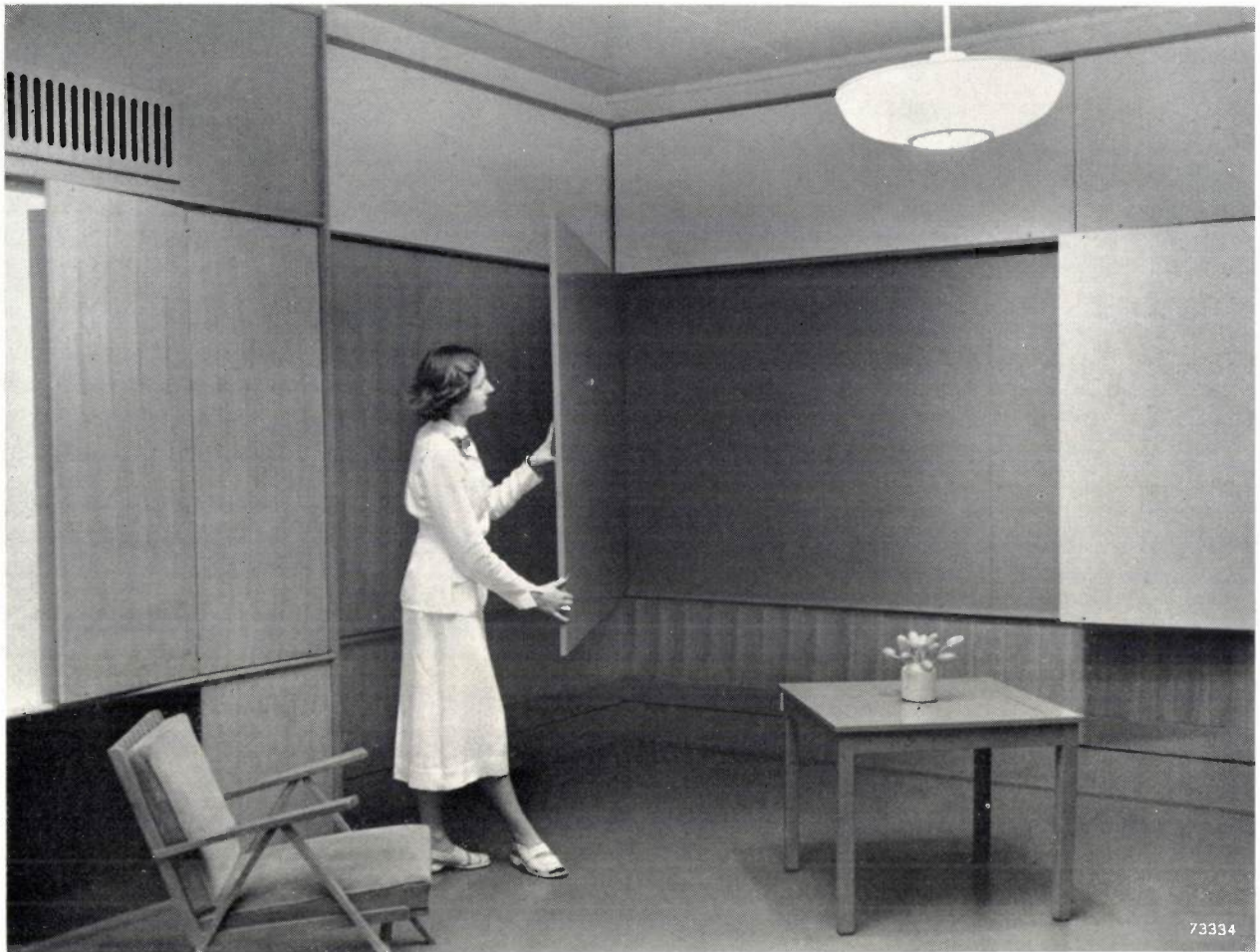
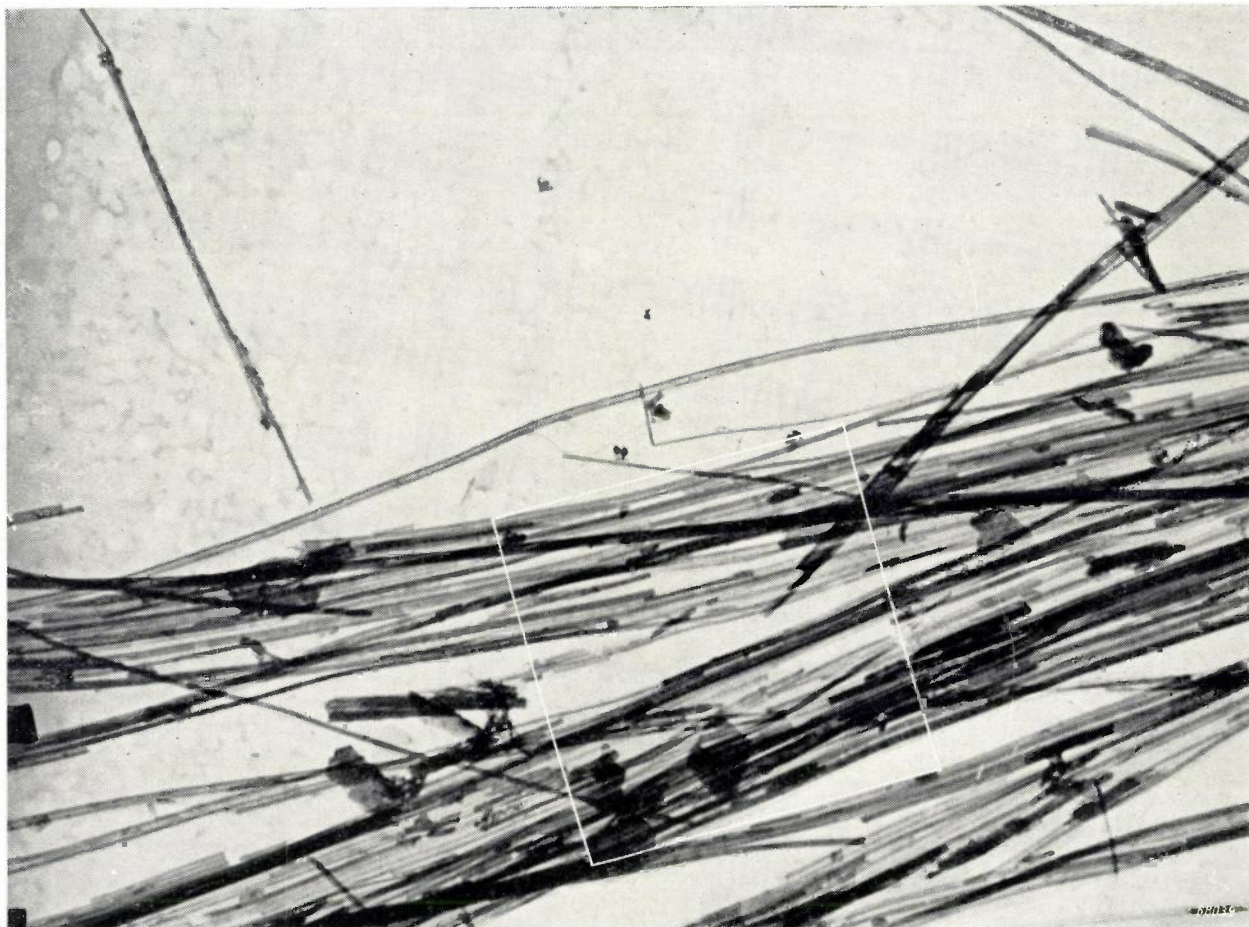


Fig. 10. Test room in the Demonstration Laboratory. Not only the lighting system but also the colours of the walls and the furniture and pictures can be easily changed.

Experience gained so far in our Demonstration Laboratory, where our theories are brought to the public and where considerable interest in them has been shown by artists, architects and interior decorators as well as by lighting experts, has been highly encouraging and has reinforced our confidence that fertile cooperation between these groups is possible.

Summary For a number of years, efforts have been made to avoid discomfort in modern lighting installations by means of

"Brightness Engineering". The author holds that instead of trying to avoid discomfort we should aim at creating comfort. This aim can only be achieved by co-operation of the lighting expert with the architect and decorator. As a starting point for such a co-operation the author proposes a scheme of lines, colours and brightnesses suitable for concentrating the attention on a visual task. The adaptation of this scheme to different situations is discussed, viz. for a single person working at a desk (for which case two centres of interest should be present), for a restaurant or a lobby (horizontal zones), for a church (vertical zones), and for a department store (a number of centres of interest). Illustrations of the practical realization of the scheme in such cases are given. Evidence for the soundness of the basic lines-colours-brightness scheme is obtained from an analysis of famous paintings and from experiments made in the Demonstration Laboratory at Eindhoven.

ASBESTOS, VIEWED THROUGH THE ELECTRON MICROSCOPE

The fibrous structure of asbestos is due to the fact that the crystals comprising the material (a magnesium silicate of variable composition) have, for the greater part, a long-drawn-out shape, frequently in the form of fine tubes. The magnification of the photograph is 30 000 \times .

The adjacent electron diffraction diagram, taken for the square indicated in the photograph, shows clearly the consequence of the preferred orientation of the group of crystals, which lie practically parallel in their longitudinal axes. The diffraction diagram was taken with the electron microscope by the method described at the time in this Review (Philips Technical Review 12, 33, February 1950).



68040

AN ELECTRICAL ROUGHNESS TESTER FOR THE WORKSHOP

by G. W. van SANTEN.

621.317.39:620.179.6

It is natural to try to test the roughness of a surface by scratching it with the nails. The same method is applied when testing the hardness of a relatively soft surface, e.g. a layer of paint. Naturally, this test is merely subjective; since this is a drawback, various attempts to find a more objective testing method were made. In the preceding volume of this review an instrument used for testing paint was discussed, which makes an impression in the layer of paint in a reproducible manner and measures its depth electrically. In the present article the author describes an instrument giving an indication of roughness measured electrically. This instrument was designed for the workshop in the first place and not especially for the laboratory, as are the expensive roughness meters already in existence.

The proper performance and the life of machine parts rubbing against each other, e.g. spindles in bearings, and pistons in cylinders, not only depend on the material, the clearance and the lubrication, but also and to a high degree on the smoothness of the surfaces moving against each other. The designer of a machine will accordingly not only specify its materials and clearances but also its roughness, or in other words, the smoothness to which the surfaces should be finished.

We can only inspect to see whether or not the latter requirement has been met if we have an instrument at our disposal that is suitable for assessing the roughness of a surface, or at any rate for comparing it with that of a sample surface. Several instruments like this are commercially available¹⁾. Most of them are very expensive instruments, however, designed for the laboratory and less suitable for the workshop. In most cases they are more accurate than necessary for the workshop and too unwieldy to handle. For this reason there is a great need for a roughness tester of a more simple and robust design, easy to handle and therefore suitable for the workshop. The electrical roughness tester discussed in the present article seems to fill this need.

A better international specification of roughness standards is no less urgently required. In the Netherlands the Central Standardization Committee has set up a Sub-Committee for Surface Conditions; this committee has concerned itself in great detail with this problem of roughness standardization. After numerous discussions with representatives of engineering works at home and abroad it has drawn up a draft standard (sheet N 785); we shall mention the main points of this presently.

¹⁾ H. Becker gives a survey: Methods of measurement and definition of surface, *Microtecnica* 4, 180-184, 276-282 and 302-307, 1950.

This draft standard has already been adopted in the Dutch Philips factories, which are dispersed across the country. Because of this dispersal a number of standard samples and derived sub-standards, in which certain numerical roughness values had been laid down, were needed to effect the adoption; the manufacturing products were compared with these sub-standards after their distribution amongst the various factories. A subjective comparison by eye or touch of nail is unreliable and of no use at all in the event of arbitration. The tester to be discussed, however, is suitable for an objective comparison.

The standard samples mentioned above were manufactured in the Philips factories at Eindhoven, by turning, grinding or otherwise machining metal surfaces until the required smoothness was obtained; this was assessed by means of various laboratory instruments. The sub-samples were derived by electro-deposition from these sample surfaces.

The Dutch Standardization Draft

The design of the Philips roughness tester is closely connected with the Dutch Standardization Draft. We shall first, therefore, discuss the chief items of this draft:

- 1) definition of the concept of roughness,
- 2) determination of a unit for the evaluation of roughness,
- 3) the laying down of roughness classes and dimension groups, i.e. subdivisions of the classes.

We shall discuss these items successively.

Definition of roughness

Fig. 1a represents the profile of a machined surface, much enlarged; the cross-section is perpendicular to the direction of machining, e.g. the grooves left by a cutting-tool on turning or planing.

The roughness of a surface such as this can be defined by four methods:

- a) the maximum difference in height between peak and valley,
- b) the difference in height between the highest peak and the mean level,
- c) the root mean square value ($h_{r.m.s.}$) of the deviation h from the mean level:

$$h_{r.m.s.} = \sqrt{\frac{1}{l} \int_0^l h^2 dx}$$

- d) the mean value h of the absolute value of h :

$$h = \frac{1}{l} \int_0^l |h| dx \quad \dots \quad (1)$$

The mean level mentioned in (b),(c) and (d) is the level resulting from a hypothetical rolling of the peaks into the valleys, giving an absolutely smooth surface. In the definitions (c) and (d) x stands for the co-ordinate along the reference line (the line of intersection of the mean level and the plane of the drawing) and l for a length great in comparison with the average distance from groove to groove. The Dutch draft is based on the mean value.

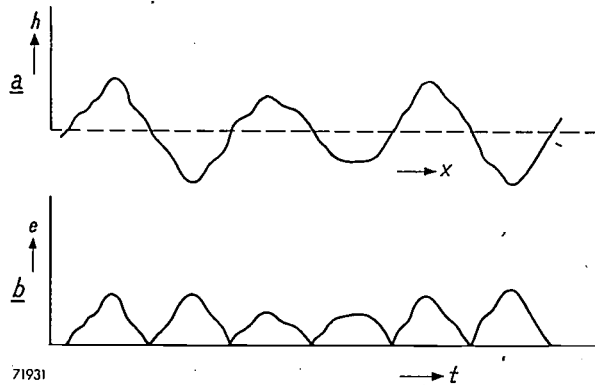


Fig. 1. a) Cross-section of a machined surface, perpendicular to the machining grooves (much further enlarged in the direction h than in the direction x).
b) Voltage e arising after full-wave rectification of the alternating voltage, excited by an electric pick-up on exploring the surface (a), as a function of the time t .

It is a short-coming of the definition (a), on which the German standard sheet DIN 4762 is based, and of the definition (b), which is advocated in France, that quite different results may be obtained from measurements made in different spots, since we may happen upon a particularly high peak or deep valley; with the definitions (c) and (d), on the other hand, a minimum measuring length l may be specified of such a size as to make the measuring result independent of l .

The American standard ASA B46 is based on the root mean square value, the British standard BS 1134 on the mean value. The Dutch committee has thought fit to conform to the latter, since measurement of the mean value is simpler and the concept can be better explained in the workshop. In practice it has been found that the difference between r.m.s. and mean values is negligible.

Roughness unit

A second important point in the Dutch draft is the introduction of a unit of roughness: the ru, a name to be taken as an abbreviation of the English "roughness unit", the French "rugosité" or the Dutch "ruwheid". The roughness of a surface amounts to 1 ru if

$$h = 1 \text{ micro-inch} = 25.4 \times 10^{-6} \text{ mm.}$$

Roughness classes and dimensions of surfaces

A system of symbols for the indication of the required roughness on technical drawings has already been adopted for a long time, e.g. $\nabla\nabla$ for what might be called "plain finished", $\nabla\nabla\nabla$ for "smooth finished", etc. These definitions are naturally vague and subjective. The most simple solution would be to replace them by a quantitative specification with a certain number of roughness units, say $\nabla\nabla = 125 \text{ ru}$, $\nabla\nabla\nabla = 32 \text{ ru}$, etc. However, an absolute classification like this would cause difficulties since different groups of technical people attach widely divergent meanings to such definitions as "plain finished", "smooth finished", etc.: a railway engine designer may call a surface smooth which is very rough to a designer of precision instruments. In analogy to the well-known ISA-system of limits and fits, the Dutch draft standard N 785 meets this difficulty by a relative classification, viz. by relating the number of ru, ascribed to the symbols, to the dimensions of the surface under discussion. This dimension ϕ stands in the case of a spindle for the diameter, in the case of a rectangular surface for the side perpendicular to the machining grooves, etc. The symbol $\nabla\nabla\nabla$ indicates a roughness of 32 ru if ϕ lies between 18 and 50 mm, but a roughness of 25 ru if ϕ lies between 6 and 18 mm, and of 40 ru for ϕ between 50 and 120 mm. Every roughness class is consequently subdivided into groups, as shown in the following table, giving a division into seven roughness classes, R_0 - R_6 ; each class is indicated by two roughness symbols, used according to the machining (see notes *) and **).

Table. Classification as to roughness, depending on the dimension σ of the surface, according to the Dutch standardization draft N 785.

Roughness class	R ₀	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆
Roughness symbols *)	∨	▽	▽▽	▽▽▽	▽▽▽▽	▽▽▽▽▽	▽▽▽▽▽▽
Roughness symbols **)	~	○	○○	○○○	○○○	●○○	●●●
Dimension σ mm	Roughness ru						
< 6	1250	315	80	20	5	1.2	0.3
6- 18	1600	400	100	25	6	1.6	0.4
18- 50	2000	500	125	32	8	2	0.5
50-120	2500	630	160	40	10	2.5	0.6
120-250	3150	800	200	50	12	3	0.8
> 250	4000	1000	250	63	16	4	1

*) Roughness symbols for surfaces with machining allowance and shaped by chip removal. (A machining allowance is an extra quantity of material on the work-piece, at those spots where machining has to be carried out.)

***) Roughness symbols for surfaces without machining allowance and shaped by chip removal, or for surfaces not shaped by chip removal.

The symbols *) hold good for e.g. a spindle that has to be turned down to a certain diameter, the symbols **) e.g. for a moulded workpiece that has to be polished. For simplicity's sake we shall only discuss the symbols *) in this paper.

Principle of the roughness tester

The roughness tester under discussion consists of a pick-up, an amplifier; a rectifier and a moving coil instrument (fig. 2). The roughness pick-up has

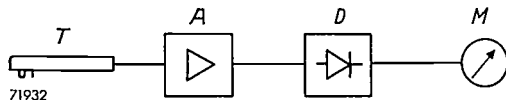


Fig. 2. Block diagram of the Philips roughness tester. T pick-up, A amplifier, D rectifier, M moving coil meter.

some analogy to a gramophone pick-up with one important difference in design, however: the former has to respond to irregularities in a direction perpendicular to the explored surface, the gramophone pick-up on the other hand to the undulations of the groove in the surface of the record.

The pick-up is moved by hand across the surface to be examined, approximately perpendicular to the direction of machining. At constant speed — we shall explain presently that this is by no means a necessary condition — it yields a voltage which is the same function of time as h is of x . This voltage is amplified by the amplifier. Rectification of the output voltage yields a direct current that is proportional to \bar{h} (fig. 1b), i.e. to the roughness to be assessed. This direct current is measured with the moving coil meter.

Construction

The roughness tester is shown in fig. 3. A few particulars are discussed below.

The pick-ups.

The cross-section of a pick-up which is used for most purposes is shown diagrammatically in fig. 4. It is of the piezo-electric type, since the smallest dimensions can be obtained in that way, which is

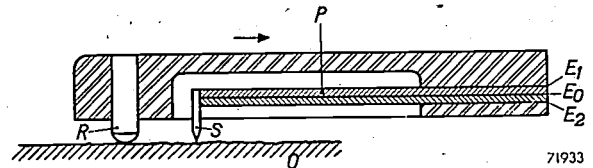


Fig. 4. Schematic longitudinal section of the pick-up, P piezo-electric element, consisting of two plates of barium titanate, and the electrodes E_1 , E_2 and E_0 , S stylus (pointed sapphire). R reference pin gliding across the tops of the irregularities of the surface O.

important for two reasons. In the first place, if the dimensions of the pick-up are small, the resonant frequencies of the moving parts lie above the vibration frequencies of the stylus when used normally, and consequently the resonances have no disturbing influence. In the second place, it is important for the dimensions of the pick-up to be small if the inner wall of small drilled holes has to be examined.

The pick-up rests with the pin R on the surface to be explored. The radius of curvature of this pin is large in proportion to the distance between the grooves, consequently the lower end of the pin remains at an approximately constant level, whilst the pick-up is moved across the surface, the reference level, in this case the plane through the peaks of the irregularities. A second pin, the stylus (S), consists of a sapphire with a radius of curvature of 60μ . This is small enough for the stylus to follow the irregularities closely, even if the groove distance is 16μ , and yet not so small that the point is easily injured or the surface scratched. The stylus is attached to one end of a piezo-electric element, the other end of this element being clamped to the body of the pick-up. As the pick-up travels along a surface, the movement of the element varies in accordance with the irregularities in the surface. Consequently, an alternating voltage is excited between the electrodes E_1 and E_2 attached to the upper and lower sides of the element. This voltage as a function of time is a copy of the irregularities. Fig. 5 is a photograph of this pick-up.

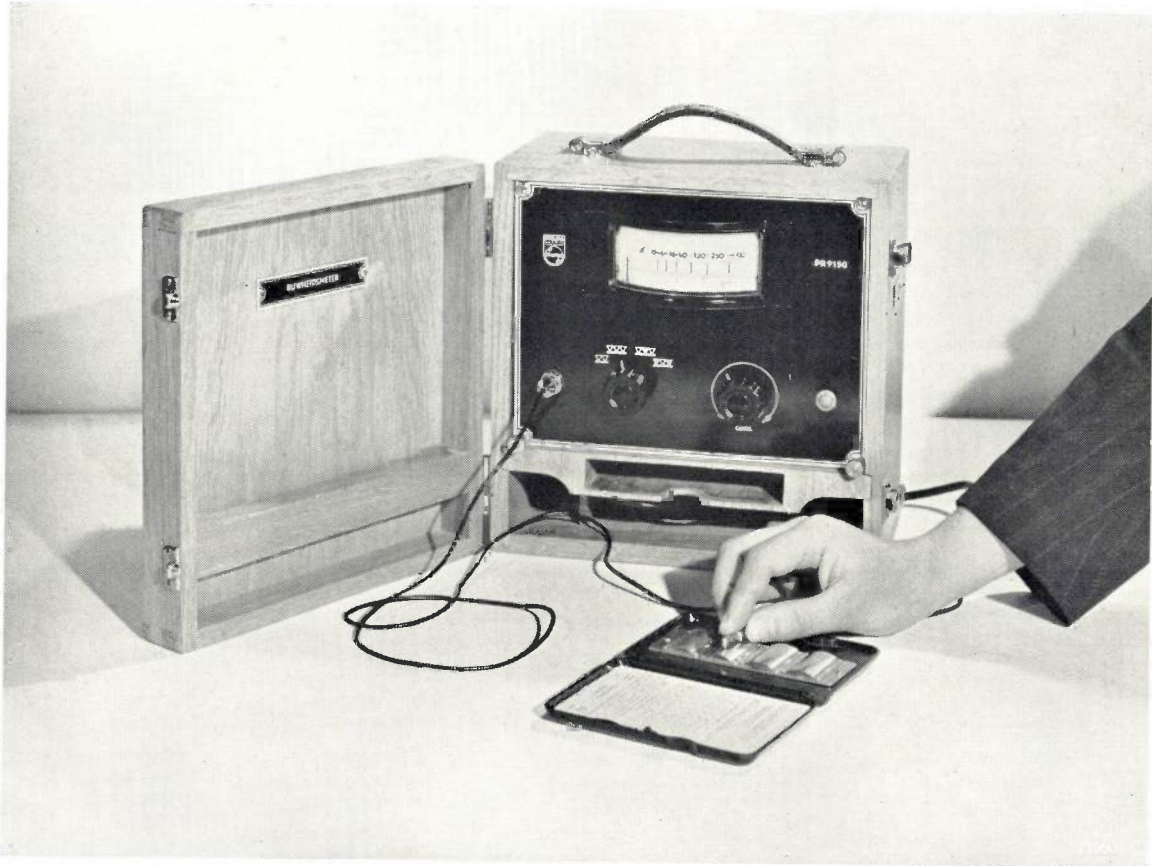


Fig. 3. The Philips roughness tester type PR 9150. Below the dial: to the left a knob for adjustment to one of the four roughness classes, to the right a correction knob. The pick-up is here traversing a sample plate of verified roughness (roughness standard); the pointer should reach to a certain calibration mark on the dial. Below the front plate is a space for the pick-ups, the flexes and the box containing the four different roughness standards (type PR 9180).

The piezo-electric element consists of two plates of barium titanate ($BaTiO_3$)²⁾, with the outer

electrodes E_1 and E_2 and a central electrode E_0 (fig. 4). The barium titanate receives its piezo-electric properties from a polarization effected during the manufacturing process by connecting a direct voltage source between the electrode E_0 and the

²⁾ J. H. van Santen and G. H. Jonker, Electrical, more especially piezo-electric, properties of barium titanate, T. Ned. Radiog. 16, 259-274, 1951 (No. 6) (in Dutch).

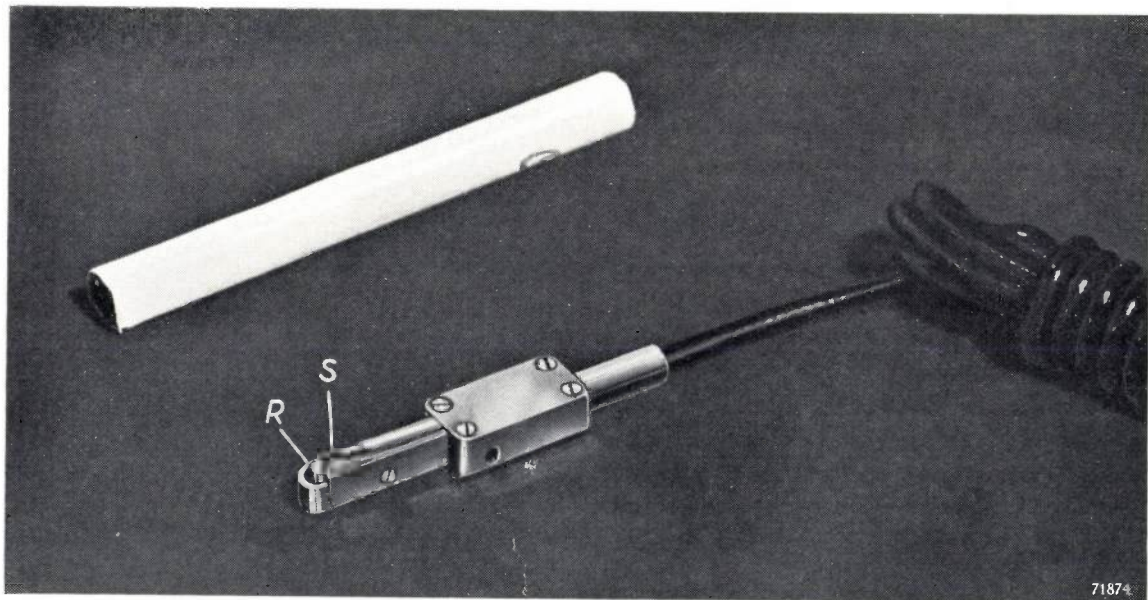


Fig. 5. The pick-up with the stylus S and the reference pin R, pointing upwards.

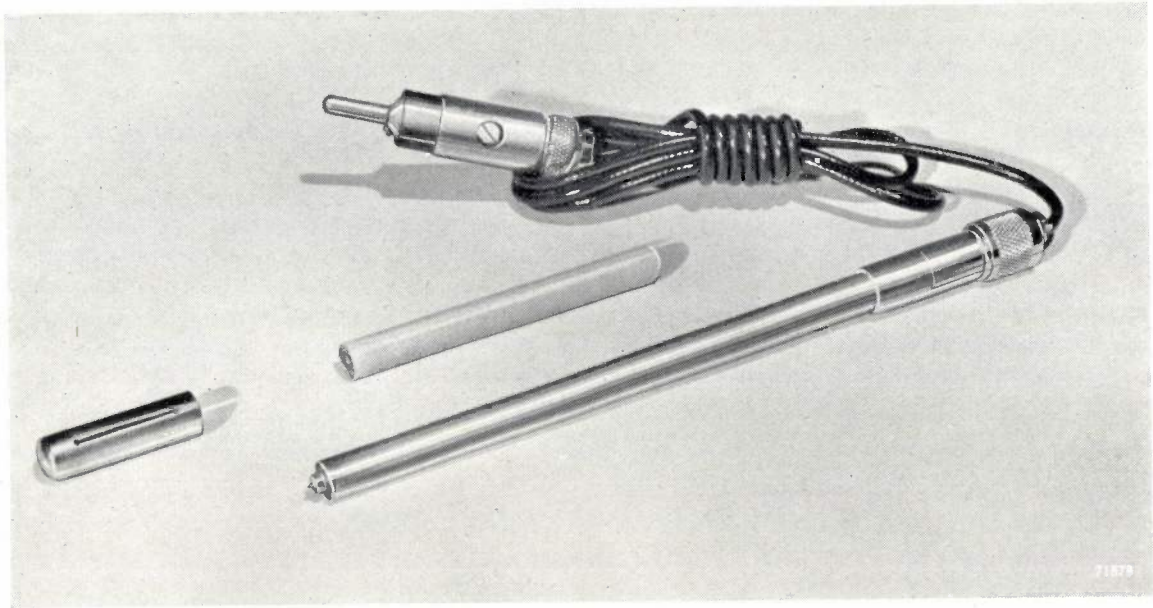


Fig. 6. Special pick-up for the examination of spots inaccessible to the pick-up shown in fig. 5.

temporarily short-circuited electrodes E_1 and E_2 ³⁾. Afterwards E_1 and E_2 are disconnected and E_0 is no longer used. Barium titanate is to be preferred to piezo-electric materials such as Rochelle-salt, because it is heat- and damp-resistant, and also to quartz because it is more sensitive and less expensive.

The height and width of this pick-up are so small (7 mm \times 4 mm) that it is still suitable for investigating the inner surface of an 8 mm hole.

The shape of the pick-up under discussion makes it unsuitable for some purposes, e.g. exploring the bottom of a so-called blind-hole (hole with flat bottom). For this reason a second pick-up was designed, rod-shaped (fig. 6), about the size of a pencil. The stylus is again a pointed sapphire, but otherwise the design differs considerably from the first pick-up. Longitudinal section (fig. 7) shows that the vertical movements of the stylus are transmitted by pieces of rubber, which exert a varying pressure on the piezo-electric element. In this case the element consists of a tube of barium titanate, with an inner and an outer coating of metal. The varying pressure excites an alternating voltage between the coatings.

The resonant frequency of this pick-up is determined by the total mass and the rigidity of the rubber. This resonant frequency should be much lower than the vibration frequencies of the stylus. The mass of the pick-up hardly follows the movement of the stylus, is this case, and the voltage

obtained is a faithful copy of the stylus movement, i.e. of the irregularities.

Since this pick-up has no reference pin, the force exerted on it by hand is completely transmitted to the workpiece by the pointed stylus. This necessitates cautious handling to avoid scratches. This danger is much smaller when the first mentioned

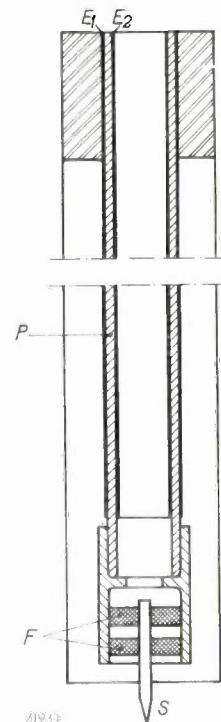


Fig. 7. Longitudinal section of the special pick-up shown in fig. 6. S stylus (pointed sapphire). F pieces of rubber. P piezo-electric element (tube of barium titanate, with electrodes E_1 and E_2).

³⁾ The fact that a permanent polarization can be induced in barium titanate, was already mentioned in Philips techn. Rev. 11, 150, 1949.

pick-up is used; whenever the shape of the workpiece is suitable this pick-up is, therefore, used for preference.

Characteristics of pick-up and amplifier

Two factors should influence the measuring results as little as possible: the speed with which the pick-up is moved and involuntary movements of the hand.

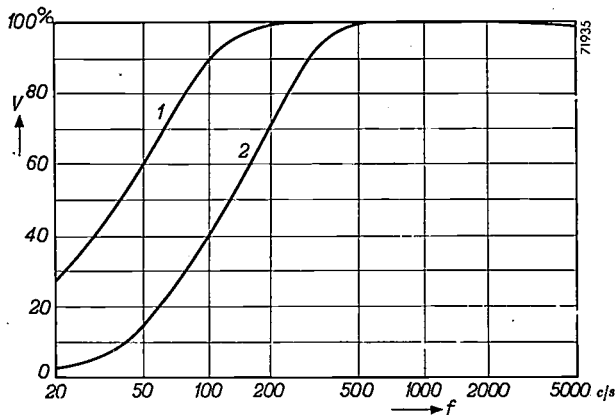


Fig. 8. Overall characteristics of pick-up and amplifier: output voltage V of the amplifier as a function of the frequency f , at constant amplitude of the sinusoidal stylus displacement. Curve 1 is used for "plain" ($\nabla\nabla$) and "smooth" ($\nabla\nabla\nabla$) machined surfaces, curve 2 for "very smooth" ($\nabla\nabla\nabla\nabla$) and "specially smooth" ($\nabla\nabla\nabla\nabla\nabla$) machined surfaces. (For the meaning of the roughness symbols, see table.)

The speed with which the pick-up is easily moved lies between about 5 and 10 cm/s. Let us suppose the maximum distance between irregularities to be 0.4 mm, which is found in case of "plain finished" surfaces; fundamental frequencies of 125 to 250 c/s will in that case arise during a normal exploring process. The involuntary movements of the hand have much lower frequencies. The frequency characteristic, i.e. the output voltage of the amplifier as a function of the frequency f , at a constant amplitude of the sinusoidally moved stylus, is shaped like curve 1 in fig. 8. It is flat for frequencies from 150 to 5000 c/s and drops steeply below 100 c/s. This shows that the undesirable voltages with very low frequencies are attenuated and that the required ones are amplified, independently of the speed with which the pick-up is moved, provided that this speed is not too low.

The speeds of a few cm/s mentioned above compare favourably with the much lower speeds of some other roughness testers. In these instruments the movement of the stylus is transmitted by means of relatively heavy parts to an element yielding a voltage. The frequency of the stylus has to be low in comparison to the already low resonant frequencies of this mechanical transmission system. For this reason the speed

of the tester has to be very low (a few mm a minute). The tester can hardly be moved by hand so slowly, and it has therefore to be driven mechanically. The measuring takes a long time, which is another disadvantage.

If the surfaces are smoother, involving smaller irregularities, the involuntary movements are relatively stronger. For this reason we change over to characteristic 2 (fig. 8), which already begins to drop at 500 c/s. This is permissible since the irregularities lie closer together if the surfaces are smoother and the vibration frequencies are consequently higher (fundamental frequencies of 1000 to 2000 c/s).

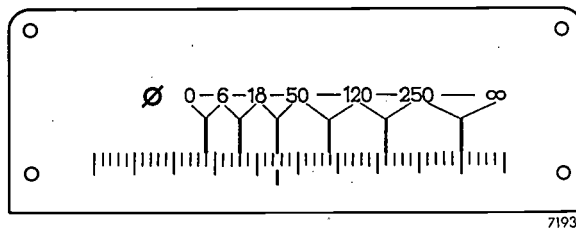
The characteristics of fig. 8 are achieved by means of filters in the amplifier.

Use of the roughness tester

The object in view when using the roughness tester is the determination of the roughness class of a surface in accordance with the specification. We have limited ourselves to the classes R_2 , R_3 , R_4 and R_5 , which occur in the vast majority of cases.

An attenuator with four positions, marked $\nabla\nabla$, $\nabla\nabla\nabla$, $\nabla\nabla\nabla\nabla$ and $\nabla\nabla\nabla\nabla\nabla$ (fig. 3), is found at the input end of the amplifier. We will assume that a surface that should have the roughness class R_4 ($\nabla\nabla\nabla$) has to be examined; we shall suppose first that the surface size falls in the group $\sigma = 18$ to 50 mm. We then proceed as follows.

The attenuator is set in the position corresponding to the roughness class expected, thus in this case to the position $\nabla\nabla\nabla$. Four calibrated roughness standards are supplied with the instrument (see below). The standard bearing the same symbol, $\nabla\nabla\nabla$, is selected; its roughness is 8 ru (the dimensions of the standards are 38 mm \times 50 mm; they fall, therefore, in the group $\sigma = 18$ to 50 mm). The degree of amplification is adjusted by means of a fine adjustment knob (fig. 3) in such a way that the meter deflects to the calibration mark belonging to $\sigma = 18$ to 50 mm, approximately in the centre of the dial (fig. 9). Thereupon the pick-up



71936

Fig. 9. Dial of the roughness tester. The thick vertical marks are reference marks, the first for surfaces with the dimension σ between 0 mm and 6 mm, the second for surfaces with σ between 6 mm and 18 mm, etc.

is moved across the surface to be examined. If the pointer of the meter is deflected more than previously, this surface is still too rough; if it is deflected less, it is already smoother than demanded.

If the dimension of the object had fallen into another group, e.g. $\sigma = 120$ to 250 μm , a roughness of 12 ru (instead of 8 ru) would, according to the table, have been indicated by $\nabla\nabla\nabla$, and the mark between the numbers 120 and 250 would have served as calibration point. Other calibration marks have been added for other groups of dimensions: $\sigma < 6$ μm , $\sigma = 6$ -18 μm , etc. (fig. 9).

Roughness standards

A box with four roughness standards (fig. 3) is supplied with the meter, each standard 38 mm \times 50 mm in size:

$\nabla\nabla$ turned	... 125 ru
$\nabla\nabla\nabla$ turned	... 32 ru
$\nabla\nabla\nabla$ turned	... 8 ru and
$\nabla\nabla\nabla$ superfinished	... 2 ru

Being chrome-plated, the standards retain their roughness calibration for years.

The inaccuracy of these roughness standards is less than $\pm 10\%$, whilst that of the measurement itself is less than $\pm 20\%$. So the error in the overall result does not exceed $\pm 30\%$. This is amply sufficient for our object, since the successive roughness classes differ from each other by a factor 4 (see table).

Summary. The roughness tester discussed above consists of two pick-ups, one amplifier, a rectifier and a moving coil meter. One of the pick-ups, which are somewhat analogous to a gramophone pick-up, is moved by hand across the surface to be examined (speed: a few cm/s). The pick-ups contain a pointed sapphire as stylus. This stylus affects a piezo-electric element of barium titanate, a material which is heat- and damp-resistant. The two pick-ups differ in form in such a way that nearly all places are accessible to one or the other. The alternating voltage obtained is amplified and rectified and the mean value of the rectified voltage is measured by means of a moving coil meter. The deflection, therefore, indicates the mean value of the irregularities, in conformity with the standardization adopted in the Dutch Philips Works. According to this standardization the roughness of a surface is 1 ru if the mean irregularity amounts to 1 micro-inch. Moreover, the connection existing between the roughness expressed in ru, the "roughness class" (indicated by symbols such as $\nabla\nabla$, $\nabla\nabla\nabla$, etc.) and one of the dimensions of the object was laid down. The dial of the roughness tester was graduated accordingly. Four roughness standards are accessories to the meter (125, 32, 8 and 2 ru).

A 3 cm-MAGNETRON FOR BEACONS

by G. A. ESPERSEN *) and B. ARFIN *).

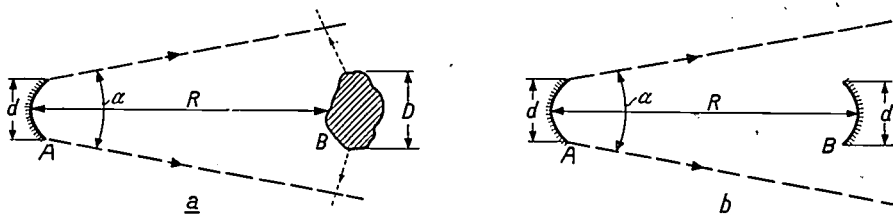
621.385.16:621.396.93

The universal acceptance of the magnetron as a pulsed oscillator for decimetric and centimetric waves is mainly due to the very high peak powers it can deliver — an essential requirement in radar, linear accelerators etc. There is, however, a need also for pulsed oscillators of comparatively low peak powers, viz, in beacon systems. This article shows how a magnetron for such a purpose has been designed under severe restrictions of weight and heater rating and how typical requirements of tuning facilities and frequency stability have been met.

Introduction

This paper will deal with a magnetron developed for beacon systems. Such a magnetron must meet rather severe requirements of frequency stability. In fact, the most important task of a beacon is to convey to distant receivers the information of its presence; for the receiving aircraft or ship it may be a matter of life or death if the absence of a signal is mistaken for an absence of the beacon. So the

available to be picked up by a receiver in the case of a beacon, or to be reflected by an obstacle in the case of radar. In the radar case, however, the signal must travel a distance R back to the transmitter again so that only a small fraction of the reflected energy will arrive at the radar receiver. The signal to be detected in this case is $8R^2/D^2$ times smaller than with the beacon, D^2 being the effective area of



73169

Fig. 1. a) Schematic drawing of a radar system. The object B of area D^2 at distance R receives a fraction $(D/aR)^2$ of the electromagnetic energy emitted by the transmitting antenna A into the solid angle α . If this energy is scattered back uniformly into half the space, a fraction $d^2/8R^2$ thereof is caught by the antenna A being at the same time the receiving antenna (diameter d). So an energy of only about $\frac{D^2}{a^2R^2} \times \frac{d^2}{8R^2}$ times the transmitted energy must be detected in the receiver.

b) Schematic drawing of a beacon system. If the receiving antenna B is of the same size (diameter d) as the transmitting antenna A , an energy of d^2/a^2R^2 times the transmitted energy is received. The signal, then, is $8R^2/D^2$ times larger than in the case of radar.

magnetron frequency must stay within the band of the receivers despite all possible changes in operating conditions (ambient temperature, output circuit impedance). Moreover, a beacon magnetron must be tunable in order to be able to avoid the frequency bands of other beacons and to compensate for the frequency spread of normal fabrication.

On the other hand, the power of a beacon magnetron can be much less than that of the better known conventional radar magnetrons. This will be clear from a glance at fig. 1a and b. At a distance R only a minute fraction of the transmitted energy is

the obstacle (the receiving antenna in both cases being supposed to be of equal size as the transmitting one). For $R = 10$ km, $D = 10$ m, this factor amounts to nearly 10^7 . Hence, whereas radar magnetrons produce powers in the order of hundreds of kilowatts, beacon magnetrons, though usually used with a smaller antenna, must supply only several dozens of watts.

At Philips Laboratories, Irvington (U.S.A.), a beacon magnetron was developed which satisfied the following conditions ¹⁾:

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

¹⁾ G. A. Espersen and B. Arfin, Low-voltage tunable X-band magnetron development, Tele-Tech, June 1951, p. 50-51 and 84, and July 1951, p. 30-31.

Frequency: nominal 9310 Mc/s, variable from 9300 to 9320 Mc/s (wavelength 3.2 cm).
 Peak power output: minimum 50 W, normal 100 W.
 Duty cycle: maximum 0.003.
 Pulling figure²⁾: 6 Mc/s.
 Frequency variation under temperature variation of $\pm 40^\circ\text{C}$: not more than ± 3 Mc/s.
 Frequency variation when duty cycle is changed: not more than ± 1 Mc/s.

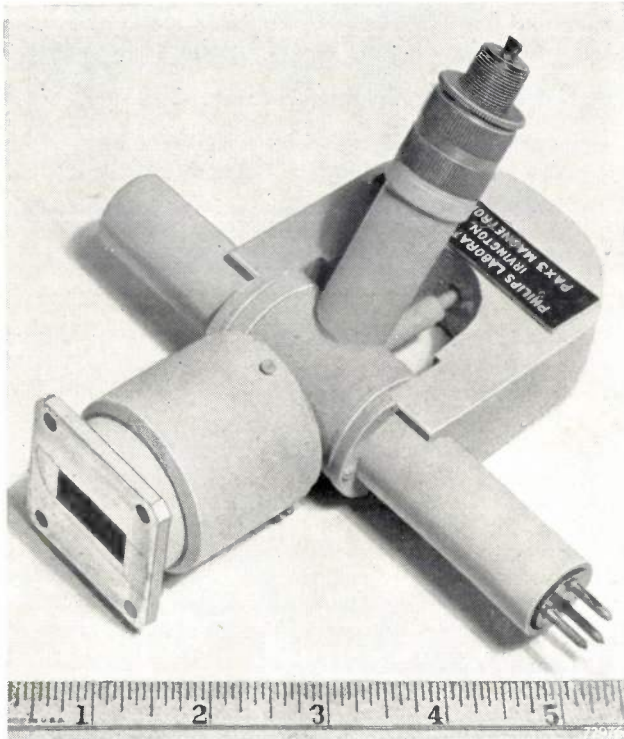


Fig. 2. Photograph of complete magnetron, type PAX3 (3 cm wavelength, 100 watts peak power), with attached magnet.

In view of the low power required, it should be possible to operate the tube with a rather low anode voltage and low cathode heater rating. Moreover, it should be possible to cut down the weight considerably, which may be of importance for portable, semi-permanent type beacons. Desirable ratings are: anode voltage 800 V, heater rating of 6.3 V, 300 mA, and total weight less than 1150 g. The efficiency should be about 30%.

Fig. 2 is a photograph of the complete magnetron (type PAX3) with its magnet; fig. 3 shows some important parts of it. The design of this magnetron will now be discussed.

²⁾ For this and other concepts used in magnetron operation and design, one is referred to: J. Verweel, *Magnetrons Philips techn. Rev.* 14, 44-58, 1952 (No. 2). This article is further quoted as I.

Type of resonators

The rising sun type of anode (see I) was chosen, comprising a number of alternately large and small resonance cavities. This anode has several advantages as compared with the strapped anode version: the internal losses are smaller; fabrication is easier, resulting in greater uniformity from tube to tube; greater mode separation is possible (as explained in I, a certain separation of the wavelengths of different modes of oscillation is necessary); finally, designs having not very different characteristics were already known in the rising sun type.

Shape and dimensions of the interaction space

The interaction space, i.e. the space where the electrons interact with the electromagnetic field (fig. 4), is defined by the number of vanes N (which, of course, is also the number of cavities), cathode radius r_c , anode radius r_a and anode height h . A parameter of minor importance in the operation of the magnetron, though of considerable interest to the tube maker, is the thickness t of the vanes.

For fixing the values of the parameters, several general relations are used:

- a) The condition of π -mode operation; in this mode the phase difference $\varphi = k \times 2\pi/N$ of the oscillations of adjoining cavities is equal to π , thus

$$k = N/2. \dots \dots \dots (1)$$

This mode is most favorable for avoiding the excitation of undesired modes (cf. I).

- b) The condition for synchronism between the rotating field and the electrons (I, eq. 5):

$$V = \frac{2\pi}{N\lambda} r_a^2 B \left\{ 1 - \left(\frac{r_c}{r_a} \right)^2 \right\} - \frac{m}{2e} \left(\frac{4\pi r_a}{N\lambda} \right)^2.$$

V is the applied anode voltage, B the flux density of the axial magnetic field, m and e the mass and charge of the electron. This equation can be written as:

$$\frac{V}{V_0} = 2 \frac{B}{B_0} - 1, \dots \dots \dots (2)$$

where

$$V_0 = 10^7 \left(\frac{2r_a}{N\lambda} \right)^2 \text{ volts}, \dots \dots \dots (3)$$

and

$$B_0 = \frac{0.043}{\left\{ 1 - \left(\frac{r_c}{r_a} \right)^2 \right\} N\lambda} \text{ Wb/m}^2, \dots \dots \dots (4)$$

λ and r_a being measured in meters. V_0 and B_0 are the coordinates of the point in the V - B -plane where the straight line defined by (2) is tangent to the cut-off parabola (cf. I, page 50).

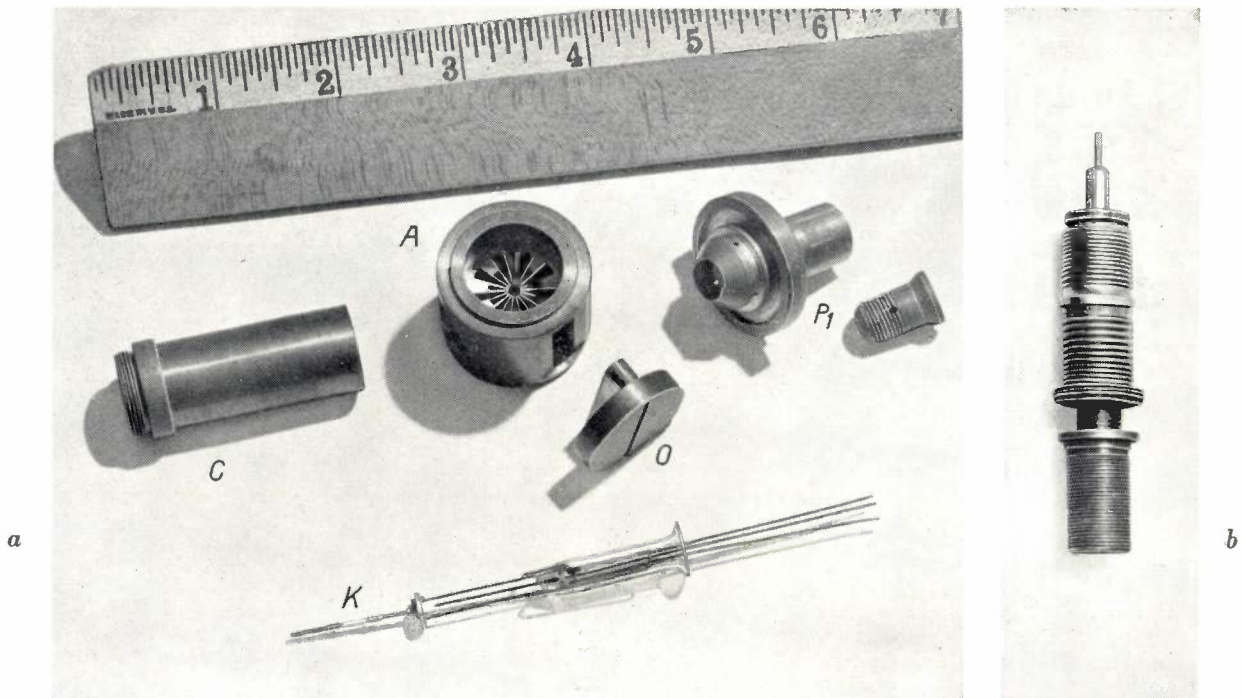


Fig. 3. a) Component parts of the magnetron shown in fig. 2. *K* cathode, *A* resonator system, *O* output transformer, *P₁* conical pole piece with core for adjustment, *C* container for tuning and temperature compensating device (see later). b) Bellows assembly for tuning and temperature compensating device (see later), on the same scale as (a).

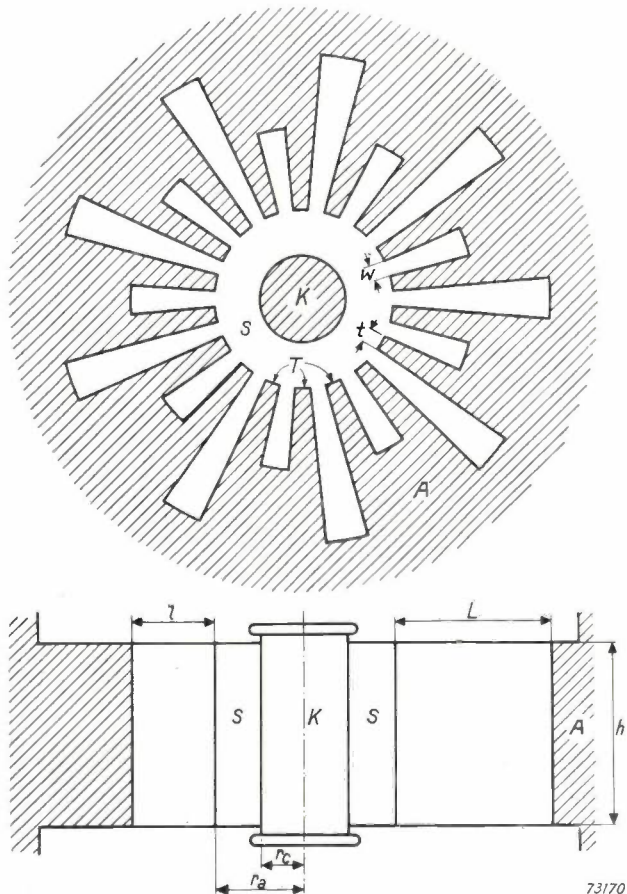


Fig. 4. Cross-section and longitudinal section of rising sun magnetron, showing the interaction spaces, cathode *K* and vanes *T* of anode *A*. *t* vane thickness, *w* slot thickness, *r_c* cathode radius, *r_a* anode radius, *h* anode height.

c) B/B_0 must be sufficiently large in order to obtain a reasonable efficiency (the efficiency is zero at $B = B_0$).

Let us first consider the number of vanes, N . As the magnet is responsible for most of the weight, we cannot afford a large B , hence also B_0 must be small. Therefore, according to eq. (4), N should be large. On the other hand, a large number of vanes means very thin vanes and tiny slots. As the anode system is produced by the hobbing technique (cf. I), both the vane thickness t and the slot thickness w — which is equal to the smallest thickness of the hob projections — should not be too small in order to avoid a mechanical failure of either the hob or the copper block during the hobbing process. A method of reconciling the conflicting requirements of large N and not too small t and w , would be to increase the anode radius r_a , but this would mean that a uniform magnetic field must be provided in a greater volume, necessitating an increase of the magnet weight. From these considerations a number of vanes $N = 18$ was chosen as a good compromise, thus making $k = 9$.

At Columbia Radiation Laboratory it was found experimentally that for 18-vane rising sun magnetrons the optimum ratio of cathode to anode radius is given by

$$\frac{r_c}{r_a} = 0,6, \dots \dots \dots (5)$$

yielding for B_0 the value (according to eq. 4):

$$B_0 = \frac{0.043}{(1-0.6^2) 18 \times 0.032} = 0.12 \text{ Wb/m}^2.$$

Now we can proceed to fix the value of B . As noticed before, efficiency increases with B , being zero for $B=B_0$. With a rising sun magnetron, however, a dip in the efficiency occurs at flux density values near $B = 0.0125/\lambda \text{ Wb/m}^2$, i.e. in our case 0.39 Wb/m^2 .

This is explained³⁾ by reference to fig. 5 which shows, at some fixed time, the currents flowing in the resonator walls of a magnetron oscillating in the π -mode. Owing to the rising sun design, the current nodes (and, hence, the places of maximum electric field strength) are not located on the ends

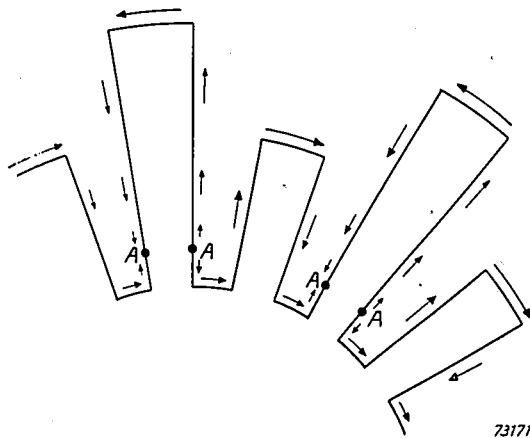


Fig. 5. The arrows indicate the direction and relative magnitude of the currents flowing at a fixed time in the walls of the cavities of a rising sun system. The high frequency current nodes occur at points A inside the larger cavities. (For the explanation of the efficiency dip at $B = 0.0125/\lambda \text{ Wb/m}^2$).

of the vanes but inside the larger cavities, at A . At any moment the currents on all vane ends evidently are flowing in the same direction and, further, the high frequency field strength across the opening of a large cavity is greater than that across the opening of a small cavity. This implies that besides the electromagnetic field characteristic of the π -mode there exists in the interaction space also a field component characteristic of the $k = 0$ mode; this is a field varying only with the radius, not with the azimuth angle. Usually this field component does not cause any harm because its effect on each electron that moves away from the cathode on a cycloid path averages zero during the time of interaction of the electron with the π -mode field, in which time the electron describes a number of loops. It will, however, be harmful when the frequency of this rotational movement of the electron, called the "cyclotron frequency" and in the case of a plane magnetron amounting to $\omega = eB/m$, coincides with the desired π -mode frequency. In fact, the periodic radial displacement of the electrons in that case will occur in synchronism with the variation of the $k=0$ field. Thus, when B attains the value

$0.0107/\lambda \text{ Wb/m}^2$, or in the case of the cylindrical magnetron $0.0125/\lambda \text{ Wb/m}^2$, λ being the desired magnetron wavelength in meters, a perturbation of the electron motion occurs, decreasing the efficiency of the interaction with the revolving wave (cf. I). Hence, in the performance chart showing the operating efficiency versus B and V surface, a "valley" will appear along the line $B = 0.0125/\lambda$ („valley of shadows").

It is, therefore, desirable to operate either above or below this flux density value. Since the magnet weight was of prime importance, it was decided to operate below the dip, at $B = 3B_0 = 0.35 \text{ Wb/m}^2$, making a good compromise between efficiency and weight.

The remaining parameters are now easily found. Eq. (2) yields

$$\frac{V_0}{V} = 5,$$

and for $V = 800 \text{ V}$, we must choose $V_0 = 160 \text{ V}$. From (3), r_a is obtained:

$$r_a = 1.15 \text{ mm},$$

and from (5) it follows that

$$r_c = 0.69 \text{ mm}.$$

The circumference of the interaction space is $2\pi r_a = 18t + 18w$. The thickness t and w of vanes and slots were chosen equal, in order to facilitate the hobbing. Hence:

$$t = w = 0.20 \text{ mm}.$$

This is a practicable value, though it will be clear that both the hob and the anode system will have to be handled with care.

The anode height h defines the necessary gap length of the magnet and the required emission density of the cathode. From experience with other magnetrons, an anode height of 5.08 mm was selected.

The cathode

In order to obtain 100 W output at 30% efficiency with 800 V anode voltage, a current I of 0.417 A is needed. The required emission density is:

$$J = \frac{I}{2\pi r_c h} = 1.9 \text{ A/cm}^2.$$

This is a reasonable value for a pulsed cathode with a duty cycle of 0.003 .

The heating power limitation is severe and a type of cathode of the highest thermal efficiency must be adopted, especially since in this low power tube heating of the cathode by back-bombardment is negligible. A normal oxide-coated cathode heated

³⁾ See J. B. Fisk, H. D. Hogstrum and P. L. Hartman, Bell Syst. Techn. J. 25, 167-348, 1946, especially p. 230.

by a single tungsten helix coated with aluminum oxide, with a rating of 6.3 V, 290 mA, meets the specifications. The rigidity and concentricity of the cathode (fig. 3, K) is ensured by supporting it axially between ceramic spacers, which are firmly seated in the recessed pole pieces.

Before proceeding to discuss the other parts of the design, it may be interesting to make a comparison between this low power tube and a high power pulsed magnetron for radar.

Let the power of such a radar magnetron of 3 cm wavelength be 1 MW. A normal efficiency value for these magnetrons is 50%. The input power, then, is $V \times I = 2$ MW. As it is known that the ratio V/I , i.e. the impedance of the magnetron, cannot be decreased to a value lower than 300-400 ohms, the required D.C. voltage and current are

$$V \approx 30 \text{ kV,}$$

$$I \approx 70 \text{ A.}$$

For 18 resonators, $r_c = 3$ mm and $r_a = 5$ mm are suitable values in this case. Calculating from these values, we obtain $V_0 = 3$ kV and $B_0 = 0.115$ Wb/m², yielding $V/V_0 = 10$ and by virtue of eq. (2):

$$B = 5.5 B_0 = 0.64 \text{ Wb/m}^2,$$

the flux density thus being above the efficiency dip.

Given the current $I = 70$ A and a practicable emission density J of 18 A/cm², the required anode height is found to be $h = 20$ mm.

As compared with our beacon magnetron, the design characteristics of the radar magnetron may be summed up as follows:

- 1) 17 times larger cathode surface and 10 times larger current density, making the total current about 170 times that of the beacon magnetron.
- 2) About 4.5 times larger anode radius r_a , entailing a higher V_0 -value, and larger ratio V/V_0 , both leading to a voltage higher by a factor of about 40.
- 3) The volume of the interaction space, and hence also the volume in which the uniform magnetic flux density must be established ($\pi r_a^2 h$), is about 75 times that of the small magnetron. Together with the larger flux density value (which results in the higher efficiency), this necessitates a total magnetic energy and a weight of magnet steel about 150 times that of the beacon magnetron!

The dimensions of the resonators and the output system

The dimensions of the resonators that remain to be fixed are the lengths L and l of the large and small cavities respectively (fig. 6).

The sum $L+l$ is determined by the desired wavelength. In fact, omitting a correction factor whose value is not far from unity, the wavelength can be put equal to the total perimeter of a large and a small cavity as indicated by the heavy line in fig. 6. The ratio L/l then must be chosen. A large value of L/l enhances the mode separation, but unfortunately also favors the $k = 0$ mode contamination (des-

cribed in the small print text on page 90), and therefore diminishes the efficiency as the B -range having a reduced efficiency is widened. Thus a compromise will be necessary.

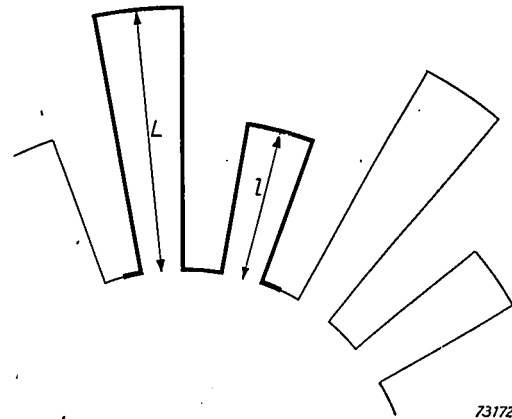


Fig. 6. The total perimeter $2L+2l$ of one large and one small cavity (heavy line) differs but little from the wavelength of the π -mode oscillation.

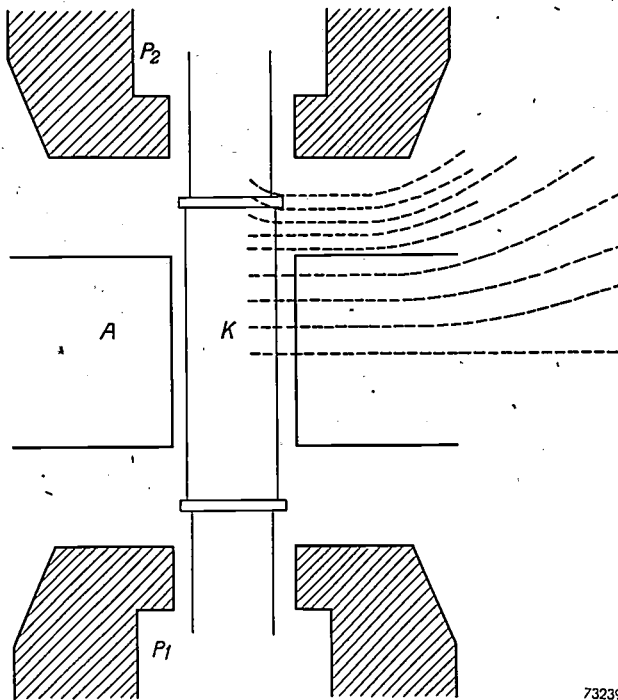
An explanation of the exact way in which the ratio L/l is selected and the correction factor for the sum $L+l$ is determined, falls beyond the scope of this article, as well as the design of the details of the output system (transformer-waveguide section): this must be designed to keep the pulling figure within the required limit of 6 Mc/s and at the same time to give the highest possible efficiency. It should only be said that the design procedure is akin to that used in the previous part of the article and to those generally applied in magnetron design: the theoretical analysis is used as far as possible, but up to the present time the theory is inadequate, i.e. it cannot predict whether or not a chosen model will operate satisfactorily: therefore, a set of dimensions is chosen from experience gained with other magnetrons of similar specifications. A model can then be built and tested, and from the result of such tests the modifications that may be necessary are derived.

In this way we obtained the value $L = 7.4$ mm and $l = 4.25$ mm, and a transforming section was developed, consisting of a rectangular slot of cross-section 22.3 mm \times 11 mm and length 10.8 mm (fig. 3, O).

The magnet

Owing to the very small interaction space, only a small magnetic energy is needed, thus providing the possibility of considerable saving in magnet weight. In designing the magnet, however, several conditions must not be neglected. The magnetic field should be essentially uniform in the entire interaction space in order to prevent electrons from spiral-

ling out to the magnet pole pieces. The vacuum gap between the pole pieces — for which it would be desirable not to exceed the anode height — must be made long enough to avoid any electrical resonances



73239

Fig. 7. Longitudinal section of the 3 cm-magnetron with pole pieces P_1 , P_2 of the magnet. Axial holes in the pole pieces permit mounting and rigid supporting of the cathode. The broken lines indicate the "equipotential surfaces" of the magnetic field, which is seen to be quite uniform in the interaction space.

of the end spaces, as these can lead to mode instability and loss in efficiency. The ultimate design consists of two conical pole pieces with a gap of 7.62 mm (fig. 3, P_1), the magnetomotive force being provided by a horseshoe magnet of weight 600 g (fig. 2). Fig. 7 is a scaled drawing of the longitudinal cross-section of the system together with the pole pieces. In this figure the "equipotential surfaces" between pole pieces as obtained from an electrolytic tank ⁴⁾ are indicated, the magnetic lines of force being perpendicular to these surfaces. It will be seen that the field is quite uniform in the entire interaction space between anode and cathode.

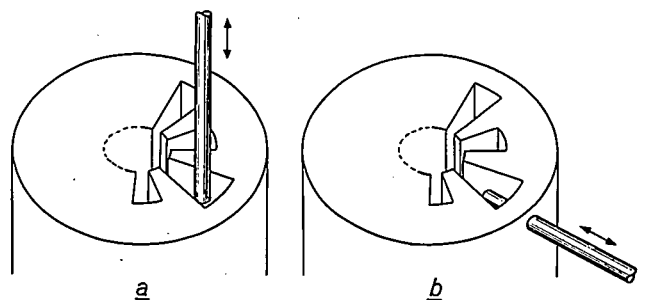
Tuning

The required tuning range, from 9300 to 9320 Mc/s, is small enough to be realized by tuning in a single anode resonator rather than in all resonators, which of course greatly simplifies the construction.

⁴⁾ M. Bowman-Manifold and F. H. Nicoll, Electrolytic field plotting trough for circularly symmetric systems, Nature 142, 39, 1938.

Two methods of tuning were investigated experimentally. In the first method a pin is introduced through one of the anode end spaces into the open side of one cavity (fig. 8a). A spurious resonance was noticed at one particular pin position, causing a marked efficiency dip and making this method unusable. In the other method, a pin is introduced radially into one cavity (fig. 8b). Experiments carried out with a non-oscillating tube indicated that the quality factor Q of the resonant system was not radically affected through a useful range of tuning. Either the small or the large resonator could be used for inserting the pin; we chose the latter.

The frequency increases as the pin is inserted; the frequency change per unit insertion and, therefore, the total frequency range varies with the cross-sectional area of the tuning pin. The frequency range obtained with three different pins and a radial motion of 2.54 mm is given in table I. The largest tuning pin, offering a range of 80 Mc/s, did not seem to affect the operating characteristics of the tube, although it decreased the efficiency slightly. In fabricating the tubes, the frequency will be found to differ slightly from tube to tube. It is, therefore, desirable to have a large tuning range so that the nominal frequency with the required range will be covered in every case. For this reason we chose the largest pin of table I.



73773

Fig. 8. Two methods of tuning the magnetron.
a) A pin is introduced through one end space into one of the open sides of a cavity.
b) A pin is introduced radially into a cavity.

Table I. Tuning range obtained with different pins and a radial motion of 2.54 mm.

Cross-section of pin	Frequency range Mc/s
Circular, 1.52 mm diam.	12
Rectangular, 1.52 mm × 3.2 mm	65
Rectangular, 1.52 mm × 4.56 mm	80

Temperature compensating device

In the temperature range through which the tube was required to operate the frequency was found to vary 0.16 Mc/s per °C due to the thermal expansion of the anode block. The permissible frequency variation is 3 Mc/s over 40°, i.e. 0.075 Mc/s per °C.

frequency by expanding the resonance cavities of the magnetron, will cause the liquid in the bellows *D* to expand comparatively strongly, while the cylinder *E* with sleeve *F* (apart from its own, smaller, expansion) maintains its position. In this case the bellows *D* will be lengthened and the pin (*B*) will

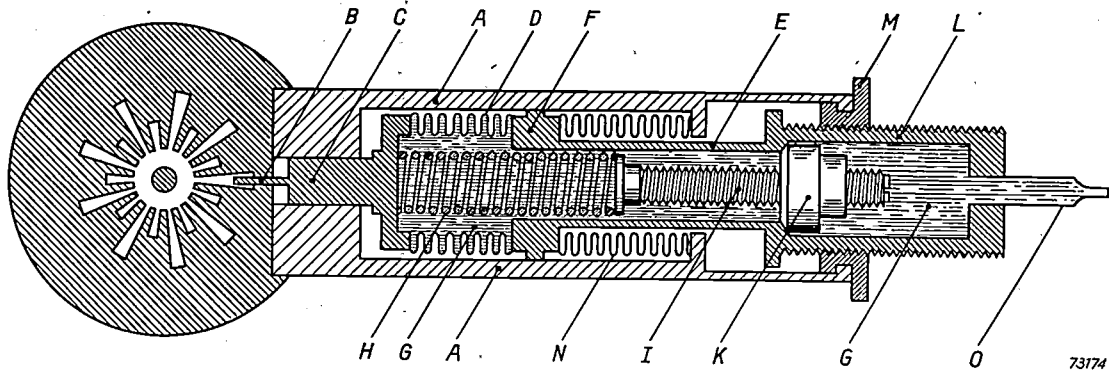


Fig. 9. Schematic drawing of tuning mechanism combined with temperature compensating device. *A* container, *B* tuning pin, *C* plunger, *D* and *N* bellows, *E* cylinder with sleeve *F*, *G* liquid, *H* pulling spring fastened to plunger *C* and pin *I*, *K* retaining nut (fluted), *L* screw, *M* knurled washer, *O* pinched-off filling tube.

It was, therefore, necessary to incorporate a temperature compensating device in the tube. This device was combined with the tuning mechanism in a novel way.

A schematical drawing illustrating the device is presented in *fig. 9* (cf. also *fig. 3b*). The mechanism is housed in a cylindrical container (*A*) brazed vacuum-tight to the anode block with its axis perpendicular to the anode axis. The tuning pin (*B*) is an integral part of a plunger (*C*) connected to vacuum-tight bellows (*D*). The other end of these bellows is soldered to the end of a cylinder (*E*) with retaining sleeve (*F*) sliding inside the container (*A*). This cylinder *E*, which is closed at its other end, and the bellows *D* are filled with a liquid (*G*) at atmospheric pressure. As the space outside the bellows *D* is evacuated, the bellows are secured against undue expansion by means of a pulling spring (*H*) fastened to the plunger (*C*) and to a pin (*I*) retained by a fluted nut (*K*).

Tuning is effected by means of a screw (*L*) and knurled washer (*M*). When the washer is turned, the whole assembly of cylinder *E* with bellows *D*, spring and plunger will move, say, to the left, thereby inserting the tuning pin in the resonant cavity and increasing the magnetron frequency. On the other hand an ambient temperature increase, which lowers the

again be inserted in the cavity, effecting a compensating increase of the frequency of the tube.

The second bellows (*N*) visible in *fig. 9* serve no other purpose than to provide in the conventional fashion a vacuum-tight seal of the cylindrical container *A* and, hence, of the magnetron; they are necessary because the sliding sleeve *F* cannot be expected to provide a vacuum-tight seal. It should be noted that on tuning, bellows *D* retain their length while bellows *N* are expanded or compressed,

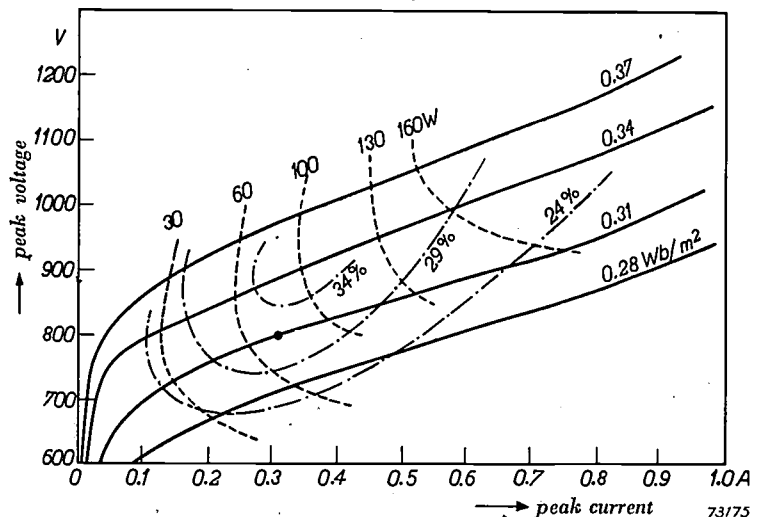


Fig. 10. Performance chart of one of the final models of the magnetron PAX3, for a duty cycle of 0.0005 (pulse width 5×10^{-7} sec, repetition rate 1000 per second). The point marked by a small circle corresponds to voltage 800 V, field 0.31 Wb/m², efficiency about 30%.

whereas on compensating temperature variations the roles of the bellows are reversed.

Filling the inner volume with liquid is effected through a copper tube (O) which then is pinched off. By selecting a liquid with a proper thermal expansion coefficient, a very satisfactory temperature compensation can be achieved. Of course, the liquid

must meet additional requirements; above all it must not chemically attack the parts with which it is in contact. For the same reason it must be absolutely free of water. Of the various liquids tried, only two proved to be satisfactory, namely, hexane and 2,2,4-trimethylpentane. The former was chosen and a temperature coefficient of the magnetron of 0.075 Mc/s per °C was obtained between the limits of -50 to +50 °C, in accordance with the specifications.

To conclude this short description, the performance chart (cf. I) of one of the final tube models is shown in fig. 10. It may be seen that the magnetron will operate satisfactorily at 800 V and at a magnetic field of 0.31 Wb/m², with an efficiency of nearly 30%. Fig. 11 is the Rieke-diagram (cf. I); the pulling figure is seen to be approximately 6 Mc/s, in accord with the specification. The total weight of this model including the magnet, is 1080 g.

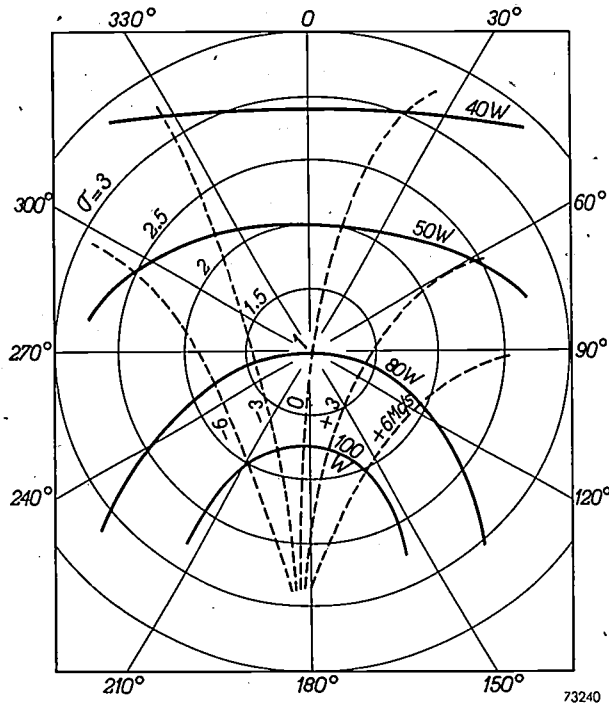
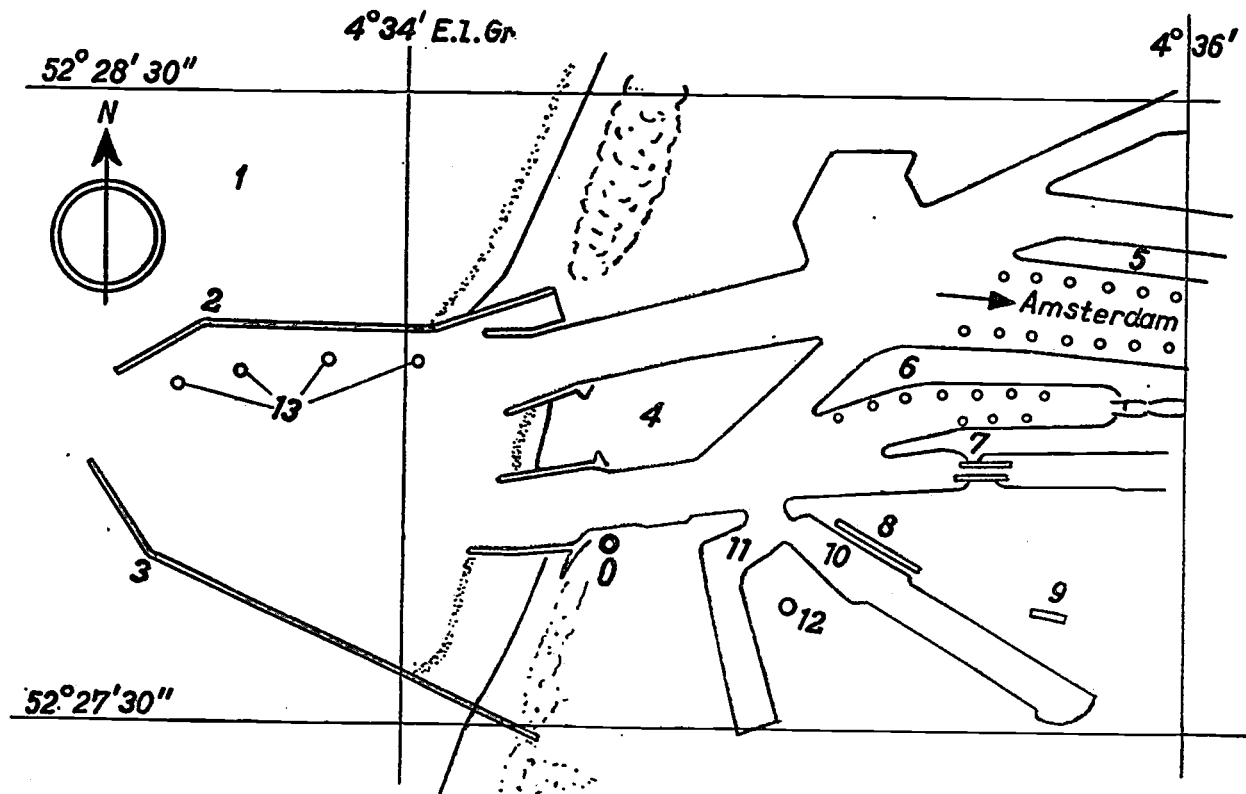


Fig. 11. Rieke-diagram of the same model as in fig. 10. Pulling figure (total frequency variation encountered on the broken lines when making one complete revolution along the circle of voltage-standing-wave-ratio $\sigma = 1.5$), is approximately 6 Mc/s.

Summary. The development described concerns an 18 vane rising sun type magnetron for 3 cm waves especially designed for beacons. The nominal peak power output is about 100 W, the frequency is variable by about 20 Mc/s, the frequency stability meets very exacting requirements, owing to the incorporation of a novel temperature compensating device. This device is combined with the tuning mechanism, based on the radial insertion of a pin into one of the 18 resonant cavities: an expanding liquid hydraulically moves the plunger carrying the tuning pin, thus compensating the temperature drift of the tube frequency. Other typical properties are: low anode voltage (800 V), low cathode heater rating (6.3 V, 300 mA) and small weight (1080 g, including the magnet). The considerations enabling a suitable design of the interaction space, the resonators, the magnet and other parts of the tube are expounded and some stress is laid on the limits of the theory and on the part played by experimental study in magnetron development.



70772

SCALE 1 : 22 500

- | | |
|--------------------------|--------------------------|
| 0 = Radar station | 7 = Southern Lock-island |
| 1 = North Sea | 8 = Fish Markets |
| 2 = North Pier | 9 = Station |
| 3 = South Pier | 10 = Fishing harbour |
| 4 = Fort Island | 11 = Herring harbour |
| 5 = Northern Lock-island | 12 = Lighthouse |
| 6 = Middle Lock-island | 13 = Buoys |

RADAR STATION AT IJMUIDEN (HOLLAND)



The 75th anniversary celebrations of the North Sea Canal, held recently, have been graced by the opening of a radar station, the first harbour radar station in continental Europe.

The installation is housed in the semaphore building at IJmuiden. The rotating aerial (a parabolic reflector) is mounted on the roof of the building. The wavelength is approximately 3 cm, and 3000 pulses are transmitted per second with a peak power of approximately 7 kW.

The photograph shows the radar panorama as it appears on the receiving screen. The photograph may be compared with the attached transparency, which is a chart of the same harbour area. The short white line just outside the piers is a ship.

When the photograph was taken, the scale of the radar image was 1 : 20 000 (that of the illustration is 1 : 22 500). Other scales may be selected, i.e. 1 : 11 000 and 1 : 57 000. In the latter case an area with about 6 miles radius is covered.

The installation was built by Philips Telecommunication Industries, Hilversum, in co-operation with N.V. Van der Heem, Voorburg (Netherlands).

LOW-HYDROGEN WELDING RODS

by J. D. FAST.

621.791.753.4

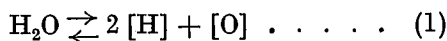
With the progress of engineering, welded joints in metal constructions have reached much higher standards. This involved the development of welding rods which yield but little oxygen to the weld metal. However, if no special precautions are taken, decrease of the undesirable oxygen content will lead to an unwanted increase in the hydrogen content. A new type of electrode, the so-called "low-hydrogen" welding rod, has therefore been developed.

Influence of the coating of the electrode on the gas content of the deposited metal

In most countries of the world only coated electrodes are used for electric arc welding, particularly for welding in air. The chief function of the coating is the protection of the molten metal from the oxygen and nitrogen in the air. Its second and nearly as important function is the evolution of sufficient gas for the globules, developed during the welding process at the end of the electrode, to be driven out with great force, giving rapid transfer of the metal to the work piece and good penetration (even in the case of "overhead" welding).

The coating of electrodes always contains metal oxides. Should these metal oxides easily yield large quantities of oxygen to the molten metal, the above-mentioned favourable function of the coating, viz. that of protecting the metal from oxygen (and nitrogen attack) would be illusory. By using very stable oxides only and by adding reducing metal powders to the coating, the latter was improved to the point of yielding a minimum quantity of oxygen to the weld metal. However the greater the success in combating the evil of oxygen absorption, the nearer becomes the incidence of another nuisance, viz. the absorption of considerable quantities of hydrogen.

The combination of compounds contained in the coating of an electrode releases water vapour on heating. This water vapour enters the arc atmosphere and is in contact with the molten metal during the welding process. Hydrogen and oxygen being soluble in this metal to a certain extent, the following equilibrium is ultimately approximated:



The square brackets indicate the hydrogen and oxygen in the metal in the dissolved state. (H and O have been used as symbols instead of H_2 and O_2 because hydrogen and oxygen are present in the metal in the atomic and not in the

molecular state). In equilibrium — depending on the temperature T — a simple relationship holds approximately, between the percentages of hydrogen and oxygen in the metal and the pressure $p(\text{H}_2\text{O})$ of the water vapour outside. This relationship is the following:

$$\frac{[\% \text{H}]^2 \times [\% \text{O}]}{p(\text{H}_2\text{O})} = C(T) \dots (2)$$

showing that the smaller the oxygen content of the metal and the larger the pressure of water vapour, the larger is the hydrogen absorption by the metal¹).

It is indeed a fact that on improvement of the electrode coatings with regard to the release of oxygen to the metal, the hydrogen absorption by the metal increased automatically at an unchanged water vapour content of the arc atmosphere. This was attended by various disagreeable phenomena.

Detrimental action of hydrogen

Hydrogen is absorbed by the molten metal during the welding process. The solubility of hydrogen in iron diminishes with decreasing temperature; at the solidification point it even drops abruptly. Thus the absorbed hydrogen will be partly released again during the cooling process. If the metal is still liquid the excess gas can leave it in two ways:

- 1) gas bubbles may form in the liquid metal and rise to the surface under the influence of the hydrostatic pressure;
- 2) the atoms may diffuse to the outer surface and recombine to form molecules.

We have already discussed extensively in this review the effect of these two processes on the welding process itself. In some circumstances, e.g. when relatively large quantities of sulphur are present in the metal, the formation of gas bubbles during the cooling process may cause porosity of the weld.

¹ The state of equilibrium to which formula (2) is applicable is never completely attained during the welding process. Nevertheless the formula will be useful for the following semi-quantitative discussion.

The article cited under 2) gives more details on the influence of sulphur.

As soon as the deposited metal has completely solidified, further release of hydrogen can only be effected via the second process: diffusion of hydrogen atoms through the metal and subsequent recombination at the surface to molecules. The thicker the metal from which the gas is to be released and the lower the temperature³⁾, the longer the diffusion process will take. Hence if the metal to be welded is very thick, a large quantity of hydrogen will be retained, in the first place owing to the long distance the atoms have to travel in the diffusion process and in the second place by the rapid cooling caused by the thickness of the metal. In this second process, as in the first one, the presence of sulphur may retard the release of hydrogen still more⁴⁾.

The excess hydrogen, left in the metal after the cooling process, can collect under very high pressure in molecular form in internal microscopic or sub-microscopic voids. The following may serve to explain this⁴⁾. Let us suppose solid iron in equilibrium with hydrogen of 1 atmosphere at 1520 °C to cool down so rapidly to room temperature that no hydrogen at all is released. At the temperature of 1520 °C about 14 cm³ hydrogen was dissolved in the atomic form in every hundred grams of iron. At room temperature less than 0,1 cm³ hydrogen is present in this form per 100 grams of iron in the equilibrium state (with hydrogen at 1 atm outside the iron). However, 14 cm³ hydrogen are actually still present in the iron after quenching and this corresponds to an external equilibrium pressure of more than 20 000 atmospheres.

The rate of diffusion of hydrogen in iron is still relatively high at room temperature; a considerable quantity of hydrogen will be released after recombination at the outer surface⁵⁾. This process of recombination to molecules will also take place at the surface of each small cavity or included piece of slag in the interior of the metal. Most important of all: at low temperatures the process does not take place at a perceptible rate in the reverse direction since only atoms can diffuse through the metal and since dissociation of hydrogen molecules at the surface is a very slow process at low temperatures. As long

as the metal is supersaturated with hydrogen the diffusion to the cavities will proceed, even if the pressure in the cavities rises to very high values. In the case under discussion this pressure can increase to thousands of atmospheres. Welding with normal types of electrode may therefore be expected to lead to the building up of large pressures of hydrogen in tiny voids within the weld. Consequently test bars made from the metal of such a weld show so-called fish-eyes on the surface of fracture (fig. 1). These fish-eyes show

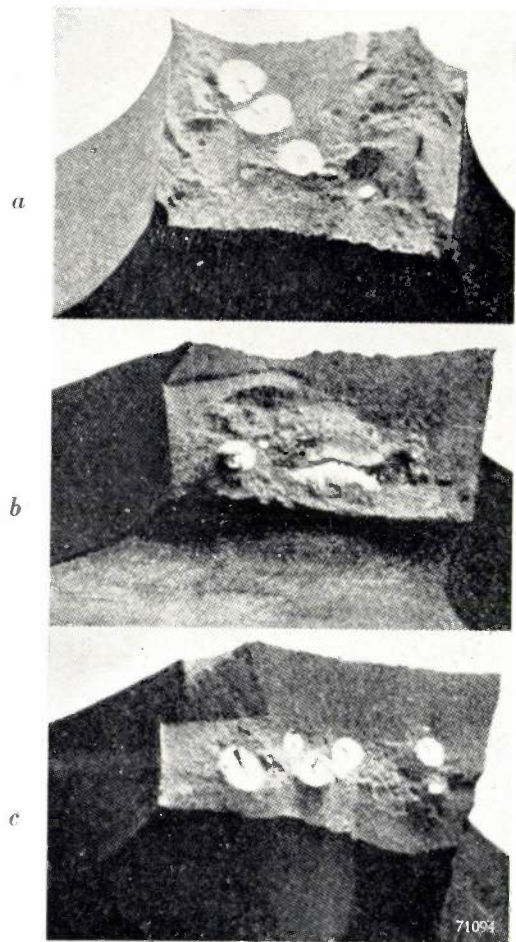


Fig. 1. Fish-eyes on the surface of fracture of test bars, made from weld metal, caused by the metal containing a) gas bubbles b) coarse slag inclusions c) finely divided inclusions. (H. Bennek and F. H. Müller, Arch. Eisenhüttenwes. 14, 605-615, 1940/41.)

minor holes or inclusions in their centre. In these regions the metal could not deform plastically because of the high hydrogen pressure and the stresses developing during the elongation. During the deformation of the test bar this results in brittle fractures on a very small scale in these regions. It was found that the pressure in the micro voids only increases to dangerously high values during the elongation. Hence fish-eyes can only

2) J. D. Fast, Causes of porosity in welds, Philips techn. Rev. 11, 101-110, 1949.

3) This holds good for iron in only one modification e.g. ferrite, for in austenite the rate of diffusion of hydrogen is smaller.

4) J.D. Fast, Hydrogen in arc welding, Lastechnik 15, 220-224, 1949 (in Dutch).

5) If a newly made weld is immersed in water, in many cases small gas bubbles will be formed at the surface for a long time.

be formed after much larger deformation than ever occurs in a construction, and therefore cannot adversely influence the strength of the welded joint.

What is much more serious, the presence of much hydrogen under very unfavourable conditions may lead to cracking in or near to the weld. This may happen with electric welding when the carbon content of the steel to be welded is very high. These cracks near the weld can be accounted for as follows: In the region of the weld the metal is heated to the austenite range (γ -iron range) during the welding process; the subsequent rapid cooling can give rise to the formation of martensite⁶⁾ (carbon-containing body-centred cubic iron with tetragonal deformation). The attendant changes in volume may in their turn cause the development of cracks of microscopic size. High hydrogen pressures may develop in these cracks unless electrodes of the low-hydrogen type are used, and these may grow into macroscopic cracks in the relatively brittle region around the weld.

Hydrogen may also contribute to cracking of the still hot metal of the cooling weld itself (so-called "hot-cracking")⁷⁾. Further discussion of this subject in the present paper would lead us too far.

Coating of some types of electrodes

It is evident from the preceding sections that the water content of the coating has increased in importance with the decrease of the oxygen transmitted by the coating to the metal. We shall go somewhat deeper into the problem by discussing the chief types of electrodes. They are respectively:

- 1) The bare electrode.
- 2) The dead-soft electrode (completely inorganic coating, chiefly consisting of iron oxides, carbonates and silicates).
- 3) The mineral-coated electrode with ferromanganese and ferro-silicon added to the coating.
- 4) The organic electrode containing i.a. large quantities of TiO_2 (rutile), although its most characteristic constituents are certain organic compounds.
- 5) The basic or low-hydrogen electrode, to be discussed later in this paper.

The two first-mentioned types of electrodes (the bare electrode and the dead-soft electrode) are hardly used any more nowadays. With the exception of the last type (the low-hydrogen electrode) all

coated electrodes contain substances, e.g. water-glass and organic compounds, which evolve relatively large quantities of water vapour during the heating process. Starting from the equilibrium formula (2), we shall try to compute the hydrogen content of the metal deposited by the various types of welding rods. To do this we shall have to know in the first place the value of the constant $C(T)$ from formula (2) for about 1540 °C, i.e. at a temperature just above the melting point of iron (1539 °C). We have pointed out in another paper⁸⁾ that at this temperature:

$$C = 1.1 \times 10^{-6},$$

if the water vapour pressure is given in atmospheres and the hydrogen and oxygen content of the metal in weight percent by. Moreover, we shall need two experimental values for each type of welding electrode, viz. the water vapour content of the arc atmosphere and the oxygen content of the liquid metal.

With regard to the latter a difficulty arises because the metal to be welded also contains silicon and manganese, which have a great affinity for oxygen and thus lower the activity of this element in the molten metal. We shall get an upper limit for the hydrogen absorption of the molten metal if we only take into account that percentage of oxygen which on chemical analysis of the deposited metal is found in the form of FeO . On the other hand a lower limit for the hydrogen absorption will be found if the percentage substituted in formula (2) is the total analytically found oxygen content. However, the latter value should never exceed the saturation content of the molten metal at 1540 °C, viz. 0.2% O. Even if more oxygen has actually been absorbed, we should never substitute a greater value since the surplus is not present in the dissolved state, whereas the quantity [% O] in the formula only refers to this state. As a matter of fact the metal absorbs more oxygen than corresponds to the solubility in molten iron if bare electrodes are used (the surplus as an oxide).

We have, basing ourselves on the experiments of Mallett and Rieppel⁹⁾, substituted in the formula a value of 0.2 atm for the partial water vapour pressure in the arc atmosphere for all coated electrodes except the low hydrogen one. For bare electrodes the water vapour content is at the utmost equal to that of the atmosphere, viz. only about

⁶⁾ For a discussion of the arrangement of the atoms in the various modifications of iron we may refer e.g. to J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips Techn. Rev. 13, 165-171, 1951 (No. 6).

⁷⁾ P. C. van der Willigen, Lasttechnik 15, 317-321, 1949; Berg- und Hüttenmännische Monatshefte 95, 234, 1950.

⁸⁾ J. D. Fast, Philips Research Reports 5, 37-45, 1950.

⁹⁾ M. W. Mallett, Welding J. 25, 396 S-399 S, 1946; M. W. Mallett and L. J. Rieppel, Welding J. 25, 748 S-759 S, 1946.

0.01 atmosphere. The same low partial pressure was entered in formula (2) for the low hydrogen electrode. We shall discuss this later.

The following table gives both the total oxygen percentage and the oxygen present as FeO for the deposited metal of the above mentioned types of electrodes. A third experimentally found content is given, viz. the hydrogen content of the deposited metal. Moreover, the table contains the lower limit of the hydrogen content calculated with the aid of formula (2) (see above).

Hydrogen and oxygen contents of the deposited metal.

Type of electrode	Experimentally found contents			Lower limit of the hydrogen content according to formula (2)	
	weight % O total	weight % O as FeO	weight % H	weight % H	cm ³ H ₂ per 100 gr. iron
bare	0.25	0.25	—	0.0002	2.5
"dead-soft"	0.20	0.20	—	0.0010	11.5
"mineral-coated"	0.12	0.04	0.0015	0.0013 ⁵	15
"organic"	0.06	0.004	0.0020	0.0019	21
"low-hydrogen"	0.03	0.001	0.0007	0.0006	6.5

Computation of an upper limit of the hydrogen content in the above-mentioned manner gives values for the two last types of electrodes that correspond to hydrogen pressures considerably higher than 1 atm. In these cases, therefore, calculation of an upper limit would have to be carried out in a slightly different manner. It would lead us too far to discuss the problem more extensively in this paper.

It will be noted that there is fair agreement between the experimentally found values and the calculated lower limits of the hydrogen content. This agreement is surprising since our calculations are but rough. We have not only based ourselves on an assumption which only holds good on approximation, viz. that a state of equilibrium is reached during welding, but — what is of more importance — we have neglected two most important factors, viz. the release of a considerable quantity of hydrogen during the solidifying and cooling process and the absorption of a large quantity of hydrogen in excess of the above calculated lower limit under the influence of silicon and manganese present in the metal. The two faults in our computation apparently counterbalance each other. As a matter of fact the experimentally found hydrogen contents exceed the figure given in the table if the cooling process is extremely rapid, whereas abnormally slow cooling results in smaller hydrogen contents.

It is evident from the table that the improvement of the quality of the weld metal by reducing the oxygen content was at first obtained at the cost of an increase of the hydrogen content. A high oxygen content is so unfavourable that bare electrodes are no longer used in most countries, although the welds produced by using these electrodes have a very low hydrogen content; dead-soft electrodes are only applied when a fine appearance of the bead is of more importance than its mechanical strength. All things considered, the two next types of electrodes, the mineral coated electrode (e.g. the Philips electrodes 50 and Contact 20) and the organic electrode e.g. the Philips 68 and Contact 18 are important improvements on the two preceding electrodes, but the hydrogen content of the deposited metal has risen to a relatively high level (especially if the organic electrode is used).

With basic electrodes, e.g. Philips electrodes 36 and 56, a weld can be made containing both little oxygen and little hydrogen. Therefore the use of these more expensive electrodes is often economically justified. We shall now proceed to a discussion of the way in which the favourable properties of these electrodes are obtained.

The low-hydrogen electrode

In principle the oxygen content of the weld metal can be lowered by choosing an electrode coating that produces a slag containing no oxides at all or only very stable oxides. The use of iron oxides should be avoided in the first place. Good results can be obtained with a mixture of stable oxides and fluorides; their proportion in the mixture should be such that the quantity of nitrogen transmitted to the metal is also small¹⁰).

For the weld to have as low a hydrogen content as possible only little water vapour should be evolved by the coating during the welding process. At first sight, one would suppose that this requirement could easily be met by heating the electrodes before the welding process to a temperature that is high enough for all the chemically bound water to be driven out. However, such a thermal treatment tends to make most electrodes unfit for use since it considerably reduces the development of gas during the welding process; we have already observed that adequate evolution of gas is of primary importance. Moreover, the heat treatment as a rule reduces the mechanical strength of the coating to such a degree that the electrodes can no longer be transported.

¹⁰) Cf. J. D. Fast, Philips techn. Rev. 10, 114-122, 1948.

However, all requirements can be met by:

a) using a coating containing a carefully chosen mixture of substances that evolve gases containing no hydrogen. b) by a well chosen thermal treatment of the electrodes before delivery. Since chiefly basic oxides are used in the coating, these low-hydrogen welding rods are often called basic electrodes.

Low-hydrogen electrodes manufactured and treated in the way described yield only very little

Use of low-hydrogen electrodes

On the practical application of non-basic electrodes it appears that their hydrogen content causes no difficulties at all in a great number of cases when normal unalloyed steel has to be welded. Low-hydrogen electrodes compare favourably, however, as soon as circumstances arise in which it is difficult for the hydrogen to escape from the metal e.g. if the steel contains large quantities of sulphur

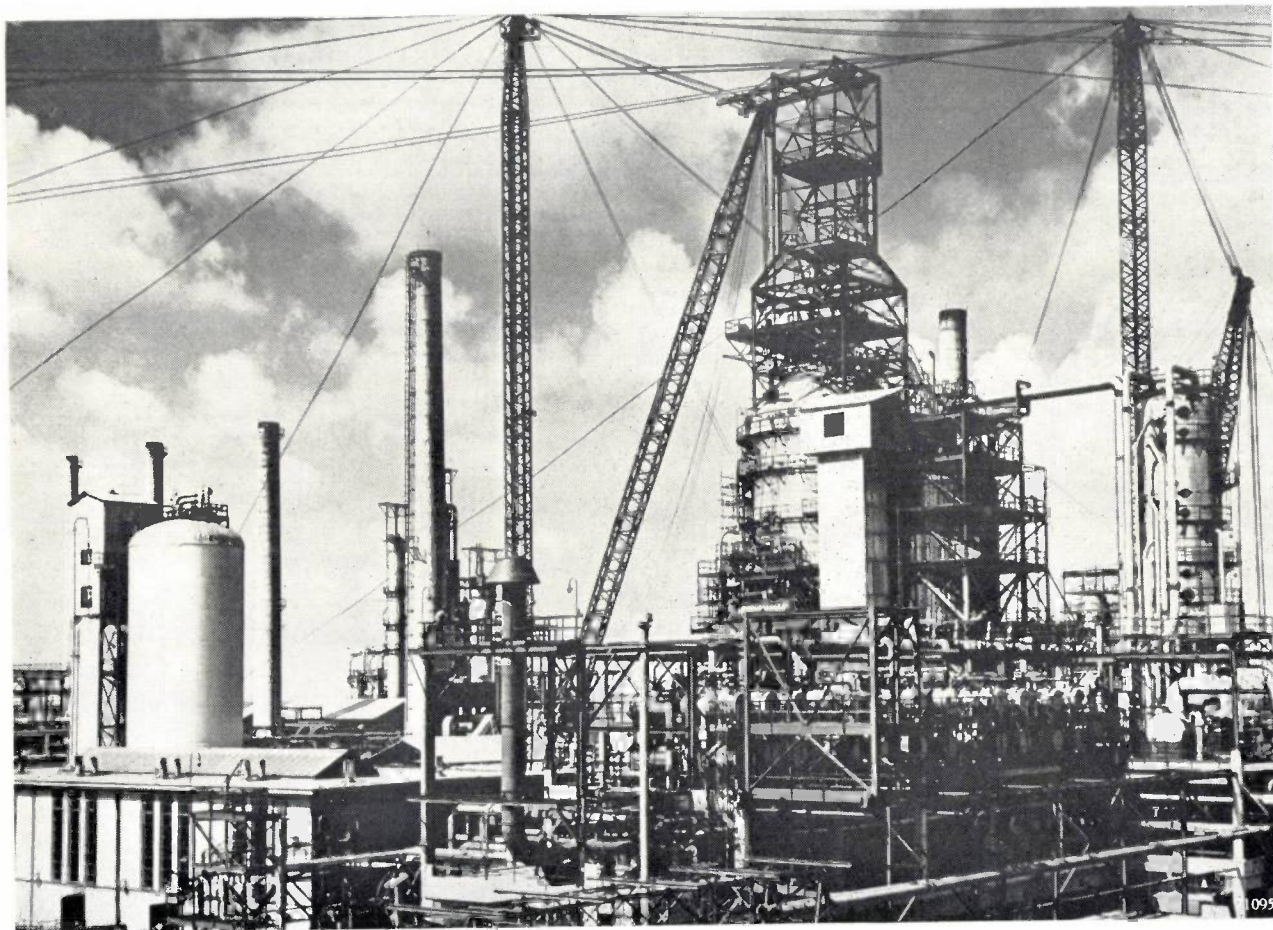


Fig. 2 A great welded project in the open air, the catalytic cracking installation of the "Bataafse Petroleum Maatschappij" at Pernis, a subsidiary of the Royal Dutch Company. This plant, with a capacity of 3500 tons high octane petrol a day, was completely welded with low-hydrogen electrodes type Philips 36. The number of electrodes used amounted to about four million.

hydrogen during the welding process; the water vapour pressure in the arc atmosphere is as low as in that of bare electrodes, viz. about 0.01 atm; according to formula (2) this corresponds to the low hydrogen content of the weld of 6.5 cm³ per 100 grams of iron, already given in the table. The various difficulties caused by hydrogen are no longer observed if these electrodes are used.

or if the material is relatively thick (although in the latter case preheating the metal may sometimes prevent dangerously high hydrogen concentrations during the welding process when using non-basic electrodes).

Care should be taken that low-hydrogen electrodes do not lose their good qualities by absorption of water vapour from the atmosphere by inadequate

storage. Zoethout¹¹⁾ tried out three simple methods to guarantee a sufficiently dry atmosphere in the storage room for low-hydrogen electrodes:

- 1) Keeping the temperature in the storage room permanently at least 10 °C higher than the outdoor temperature.
- 2) Placing containers filled with hygroscopic substances in the storage room.
- 3) Continuously conveying dried and heated air into the storage place.

In practice each of these methods proved to be sufficient.

When welding in the open air, a storage place that comes up to one of the three above-mentioned requirements will in most cases not be available. The following procedure was therefore adopted when the large project shown in *fig. 2* had to be carried out in the Netherlands. The basic electrodes were delivered in boxes by the Philips welding rod

factory to the unheated storeroom; these boxes were at once placed in tins which were sealed with gummed tape. Packed in this way the electrodes could be stored without any deterioration in quality. Naturally, after a box was opened, some of the electrodes were in contact with the open air for several hours. However, the moisture absorbed by the welding rods during this period was never sufficient, not even when there was a thick fog, for poor hydrogen-characteristics to develop during the welding process.

Summary. Nowadays coated welding rods are almost exclusively used for electric arc welding. The coating should transmit little oxygen (and nitrogen) to the metal during the welding process, since otherwise bad quality welds are obtained. Suppression of the oxygen transmission causes absorption of large quantities of hydrogen if non-basic welding rods are used. In unfavourable circumstances this may give rise to porosity and cracking of the weld. In order to decrease the hydrogen content, the welding rod should contain as little water as possible. Basic or low-hydrogen electrodes were developed to meet this requirement. For these electrodes to keep their favourable properties, they have to be stored in a special way.

¹¹⁾ G. Zoethout, *Lastechniek* 17, 126-128, 1951.

AN INSTRUMENT FOR MEASURING COMPLEX VOLTAGE RATIOS IN THE FREQUENCY RANGE 1-100 Mc/s

by G. THIRUP.

621.317.34: 621.396.645

In the early days of amplifier technique, flatness of the gain-frequency curve over the audio frequency range was considered a reliable criterion of the satisfactory transmission of speech and music. Later, however, with the advent of pulse and television transmission, it was found that a flat gain-frequency characteristic was insufficient to ensure conservation of the waveform of the signal. Moreover, the present wide-spread use of negative feedback in amplifiers necessitates a convenient method of investigating the stability of such amplifiers. The measurement of the complex ratio of two voltages is an essential step in both these techniques.

The instrument described, with which such ratios can readily be measured, was designed specially for the investigation of amplifier stability, over a much wider frequency range than has hitherto been possible. It is intended for research work in the Philips Laboratories and is not commercially available.

A communication system generally consists of a signal source, a transmission system and a load represented by a terminating impedance. The transmission system — assumed to be linear — is characterized by the ratio between an applied sinusoidal input voltage and the corresponding output voltage. This ratio is, in general, complex and a function of frequency. The numerical value, or modulus, of the ratio is easily measured by means of voltmeters or by inserting a calibrated variable attenuator and reducing the ratio to a datum value. For many purposes a gain-frequency characteristic measured in one of these ways is sufficient information. In other cases, however, it is essential to know the phase difference between the input and output voltages, as a function of frequency.

The necessary and sufficient conditions to ensure the conservation of the waveform of an arbitrary input signal, transmitted, for example, by a video amplifier, are:

- 1) that the amplitude characteristic be flat over the frequency range involved ¹⁾,
- 2) that all frequencies present be delayed by the same amount.

The latter condition means that the phase difference between the sinusoidal input and output voltages must be proportional to the frequency; any number of complete phase reversals, may, however, be introduced without affecting the wave-form.

In a communication system with a certain bandwidth (say, 4 Mc/s), it is sufficient in practice to know the amplitude and phase characteristics throughout this band and only a little beyond both limits. For a certain link in such systems, however, a much wider range of measuring frequencies is indispensable. The case we have in mind is that of the feedback amplifier. The amplifier itself and the feedback network together form a closed circuit. If this circuit is interrupted at any point, the ratio of the voltages at the output and input terminals so obtained can be measured and plotted in the complex plane; the result is the Nyquist diagram, which yields information about the stability of the amplifier ²⁾. Now this Nyquist curve has to be extended over a frequency range that on both sides widely exceeds the band for which the amplifier is designed — in fact, from a frequency some 0.1 to 0.05 times the lower limit of the band, up to a frequency some 10 to 20 times the upper limit (*fig. 1*). Measurements at the lower end of this frequency range generally present no difficulties. Often they can even be omitted and the information obtained by calculation, as at rather low frequencies the network constants in most cases are known with sufficient accuracy. At the upper end of the range, however, stray capacitances and transit time effects must be taken into account; this makes calculation too involved and unreliable, if not impossible. Therefore,

¹⁾ More strictly speaking: it should have the shape of a Gaussian curve. See, for instance, G. E. Valley and H. Wallman, Vacuum tube amplifiers, Appendix A, Radiation Lab. Series 18, McGraw Hill, New York 1948.

²⁾ H. Nyquist, Regeneration theory, Bell Syst. techn. J. 11, 126-147, 1932.
B. D. H. Tellegen, Philips techn. Rev. 2, 292, 1947.
H. W. Bode, Network analysis and feedback amplifier design (chapter 8), D. van Nostrand, New York 1945.
F. Strecker, Die elektrische Selbsterregung, Hirzel, Stuttgart 1947.

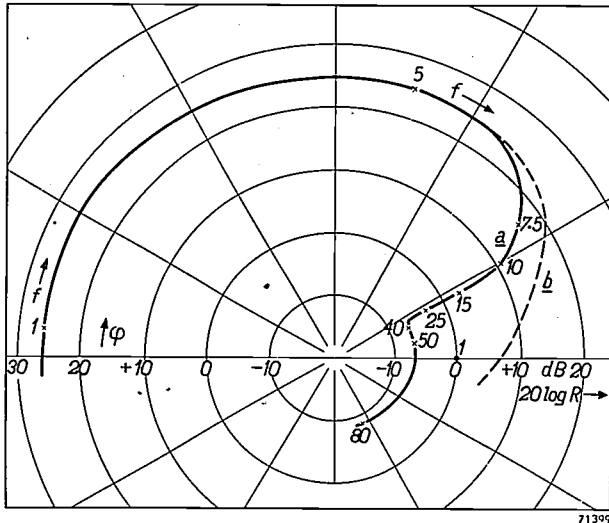


Fig. 1. Complex voltage ratio of two feedback amplifiers, plotted in the complex plane. According to the Nyquist criterion, the amplifier with curve *a*, not comprising the point 1, is stable, and the other, with curve *b*, embracing the point 1, is unstable.

The curves show the locus of the voltage ratio, with the frequency *f* as a parameter. The amplifiers are intended, for example, to operate over a frequency range of 1.5 Mc/s. In order to obtain the decisive part of the Nyquist curve, measurements have to be carried out at frequencies much higher than the upper limit of this band.

measurements are in this case essential. The instrument to be described was designed for the determination of the Nyquist curve at the higher frequencies. Its frequency range extends from 1 to 100 Mc/s.

General description and theoretical considerations

Choice of the measuring method

An instrument for measuring complex voltage ratios must satisfy certain practical requirements in order to render it suitable for development work. It is highly desirable, for instance, that the change from one measuring frequency to another can be accomplished by means of one knob only, and that the complex ratio can be read directly from dials, either in the form $x + jy$ or as a modulus *R* and phase angle φ . The accuracy must be better than 0.2 dB in *R* and 2° in φ . Further, the input terminals of the instrument must be mechanically capable of being introduced into the amplifier circuit at the points at which the voltages are to be compared, in order to avoid long connections between these points and the instrument. This means that the input terminals have to be made in the form of voltage pick-up devices, or probes, attached to the instrument itself by means of flexible cables. The input impedance of these probes must be so high as to form a negligible

load on the circuit at the points with which they are brought into contact.

Various methods for measuring complex voltage ratios have been investigated by Lévy³). The necessary accuracy can best be obtained by a null method, as this requires no voltmeter readings, which are often an important source of errors. A simple and accurate null method, however, is only feasible at a fixed and rather low frequency. This does not necessarily restrict the measuring frequency. By modulating the measuring frequency f_m with an auxiliary frequency f_a , the former may be converted into a fixed intermediate frequency $f_i = f_a - f_m$ below the range of frequencies required for the measurements⁴). The frequency conversion preserves the original complex voltage ratio.

Let $v_1 = V_1 \cos \omega_m t$ denote the input voltage and $v_2 = V_2 \cos (\omega_m t + \varphi)$ the output voltage of the network under test, ω_m being $2\pi f_m$. Then V_2/V_1 is the amplitude ratio and φ the phase difference to be measured.

Let $v_3 = V_3 \cos (\omega_a t + \psi)$ denote an auxiliary voltage of frequency $f_a = \omega_a/2\pi > f_m$; ψ is an arbitrary phase angle. The intermodulation products of v_1 and v_3 include a voltage u_1 :

$$u_1 = V_1 f_{13}(V_3) \cos (\omega_i t - \psi).$$

The corresponding intermodulation product u_2 , due to mixing v_2 and v_3 , is given by the equation:

$$u_2 = V_2 f_{23}(V_3) \cos (\omega_i t + \varphi - \psi).$$

Here $\omega_i = \omega_a - \omega_m$. As is apparent, u_2 and u_1 have the same phase difference φ and — provided the two mixers are similar, i.e. $f_{13}(V_3) = f_{23}(V_3)$ — the same amplitude ratio as the original voltages v_2 and v_1 . The frequency $f_i = \omega_i/2\pi$ of u_1 and u_2 , however, can be made much lower than f_m .

Due to this preservation of the complex voltage ratio, a measurement at a fixed and intermediate frequency can be substituted for the measurement at a variable and very high frequency f_m , compensation being obtained at a fixed, much lower frequency f_i . This, in fact, is the principle adopted in the instrument to be described.

The frequency-conversion system requires two generators, one supplying the variable measuring frequency f_m , the other the auxiliary frequency f_a differing from f_m by a fixed amount f_i . Finding a practical solution for the construction of these generators, with the necessary high constancy in frequency difference f_i , was one of the main problems. Another was the avoidance of stray effects that would falsify the measurements; questions relating to this problem are treated in an Appendix.

³) M. Lévy, Methods and apparatus for measuring phase distortion, *Electr. Communication* 18, 206-228, 1940.
⁴) E. Peterson, J. G. Kreer and L. A. Ware, *Bell Syst. techn. J.* 13, 680-700, 1934.

In the third place we mention the design of a novel phase shifting device used for balancing the intermediate-frequency voltages.

The instrument can be divided logically into a high-frequency circuit comprising the generators, the mixers etc., and an intermediate-frequency (and associated low-frequency) circuit consisting of the balancing and indicating devices.

They are kept constant in frequency by crystal oscillators Q_1 and Q_2 , generating the frequencies 200.3/27 and 200.0/27 Mc/s, respectively. Both these frequencies are multiplied 27 times by three successive triplers (see blocks Mu_1 , Mu_2). Another oscillator, O_v , has a frequency f_v , variable from 201.3 to 300.3 Mc/s. Fractions of the voltage generated by O_v are applied to the oscillators O_1 and O_2 , which at the same time function as mixers.

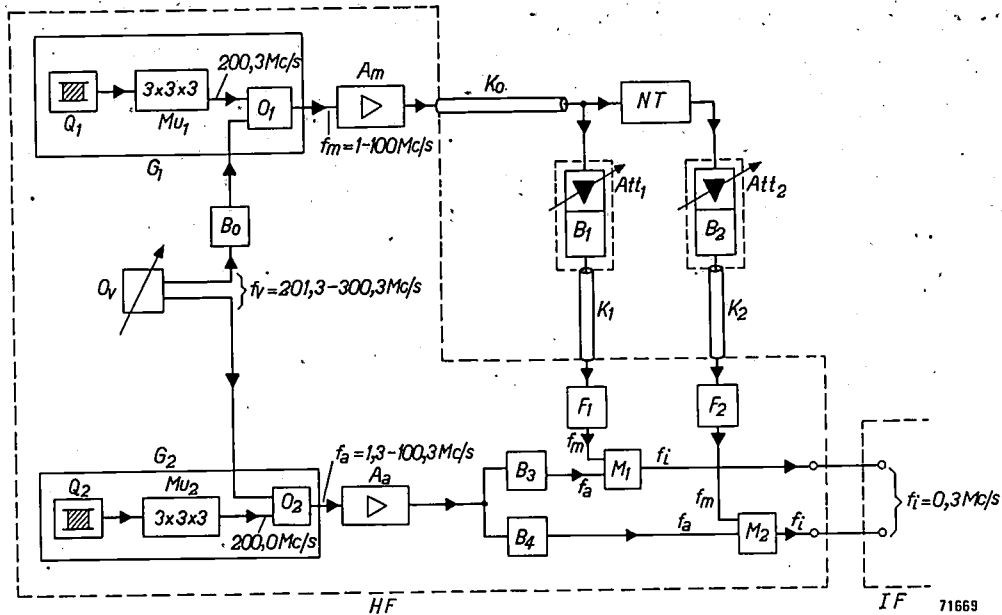


Fig. 2. Block diagram of the H.F. circuit. G_1 generator comprising a crystal oscillator Q_1 and a frequency multiplier Mu_1 , synchronizing an oscillator O_1 at the frequency 200.3 Mc/s. G_2 generator similar to G_1 but with a frequency 200.0 Mc/s. O_v oscillator with a frequency f_v variable from 201.3 to 300.3 Mc/s. The measuring frequency $f_m = f_v - 200.3$ Mc/s. = 1 to 100 Mc/s is obtained from O_1 , working as a mixer. An auxiliary frequency $f_a = f_v - 200.0$ Mc/s is similarly obtained from O_2 . $f_i = f_a - f_m =$ intermediate frequency = 0.3 Mc/s. A_m , A_a wide-band amplifiers. K_0 , K_2 , K_1 coaxial cables. NT network under test. Att_1 , Att_2 variable attenuators. B_0 , B_1 , B_2 , B_3 , B_4 buffers. F_1 , F_2 band-stop filters. M_1 , M_2 mixing circuits. HF denotes the H.F. chassis, IF the I.F. chassis.

The high-frequency circuit

The production of two frequencies differing by an approximately constant amount can be accomplished by mechanically ganging the tuning capacitors of two valve oscillators; a similar technique is adopted in normal superheterodyne receiving sets, in which the tuning circuits of the H.F. amplifier and the local oscillator are ganged together. In our case, however, where $f_i = 300$ kc/s and the maximum tolerable deviation is only 0.5 kc/s, mechanical ganging would not provide the necessary constancy in frequency difference.

The solution adopted is shown in principle in fig. 2. The generators G_1 and G_2 provide the fixed frequencies 200.3 and 200.0 Mc/s, respectively.

One of the intermodulation products of O_1 has the frequency $(f_v - 200.3)$ Mc/s, which is used as the measuring frequency f_m . Since f_v is variable from 201.3 to 300.3 Mc/s, f_m has a range from 1 to 100 Mc/s. In the same way the auxiliary frequency $f_a = (f_v - 200.0)$ Mc/s is obtained from O_2 . The frequency f_a differs from f_m by the fixed amount of 0.3 Mc/s, i.e. by the intermediate frequency f_i .

The output voltages of G_1 and G_2 are amplified in the wide-band amplifiers A_m and A_a , respectively. A cable conveys the output of A_m to the network under test. The input and output voltages of this network are applied, via attenuators (Att_1 , Att_2), to mixing circuits (M_1 , M_2), where they are mixed with voltages derived from the amplifier A_a .

Interposed buffer stages (B_1, B_2, B_3, B_4) prevent unwanted couplings (see Appendix). The mixers produce intermediate-frequency output voltages, which are compared in the intermediate-frequency circuit to be described in the next paragraph.

The intermediate-frequency circuit

As shown above, the high-frequency circuit delivers two 0.3 Mc/s output voltages having the same complex ratio as the high-frequency voltages present at the input and output terminals, respectively, of the network under test. In the intermediate-frequency circuit, these 0.3 Mc/s voltages are made to cancel each other by changing the amplitude and phase of one of them until the sum of the two voltages becomes zero. The balance indication is obtained in the following way. The voltages to be compared are applied to the input terminals of an amplifier (A_{IF} , fig. 3). With the aid of a beat oscillator the out-of-balance voltage is converted into a low-frequency voltage, which is made audible by means of an audio-frequency amplifier and a loudspeaker. The vanishing of the beat note indicates balance. The amplitude and phase adjustments necessary for the above operations are accomplished by means of a variable attenuator Att_3 and a variable phase shifter ϕ (fig. 3).

From the dials of these instruments the ratio can be read directly as an amplitude change in dB and a phase shift ⁵⁾.

Audible indication is to be preferred to visual, as it has the considerable advantages that beat notes due to unwanted intermodulation products are easily distinguishable from the normal one, and that the influence of the noise level is smaller.

On the input side of the IF circuit, similar filters (F_3 and F_4) pass the intermediate frequency

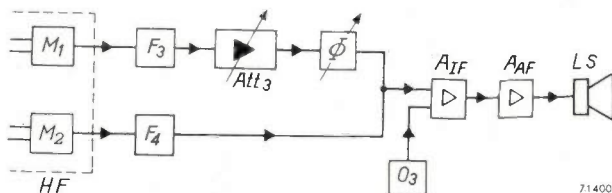


Fig. 3. Block diagram of the I.F. chassis. F_3, F_4 low-pass filters. Att_3 variable attenuator. ϕ phase shifter. A_{IF} I.F. amplifier. O_3 beat oscillator. A_{AF} A.F. amplifier. LS loudspeaker. M_1, M_2 , mixing circuits (see fig. 2).

Att_3 and ϕ are so adjusted that the beat note from the loudspeaker vanishes, indicating that the two I.F. voltages applied to A_{IF} balance.

⁵⁾ Attenuators Att_1 and Att_2 were here supposed to have equal attenuations. If they have different settings — as in general will be the case — the difference in readings of Att_1 and Att_2 must be added to the reading of Att_3 .



Fig. 4. View of the complete instrument. From top to bottom: loudspeaker panel, I.F. panel, H.F. panel, table (with the two probes) and supply panel.

f_i and stop the high frequencies f_m and f_a that are always present to some extent at the output terminals of the mixers.

The amplifier A_{IF} is selective. It need not satisfy special requirements in respect of phase or amplitude, since its only function is to amplify the out-of-balance voltage.

The attenuator Att_3 must not introduce any change of phase shift over the entire attenuation range (the same applies, by the way, to the attenuators Att_1 and Att_2 of the H.F. circuits, fig. 2); the phase shifter must similarly have an attenuation that is independent of the setting. Details of these components will be discussed in later paragraphs.

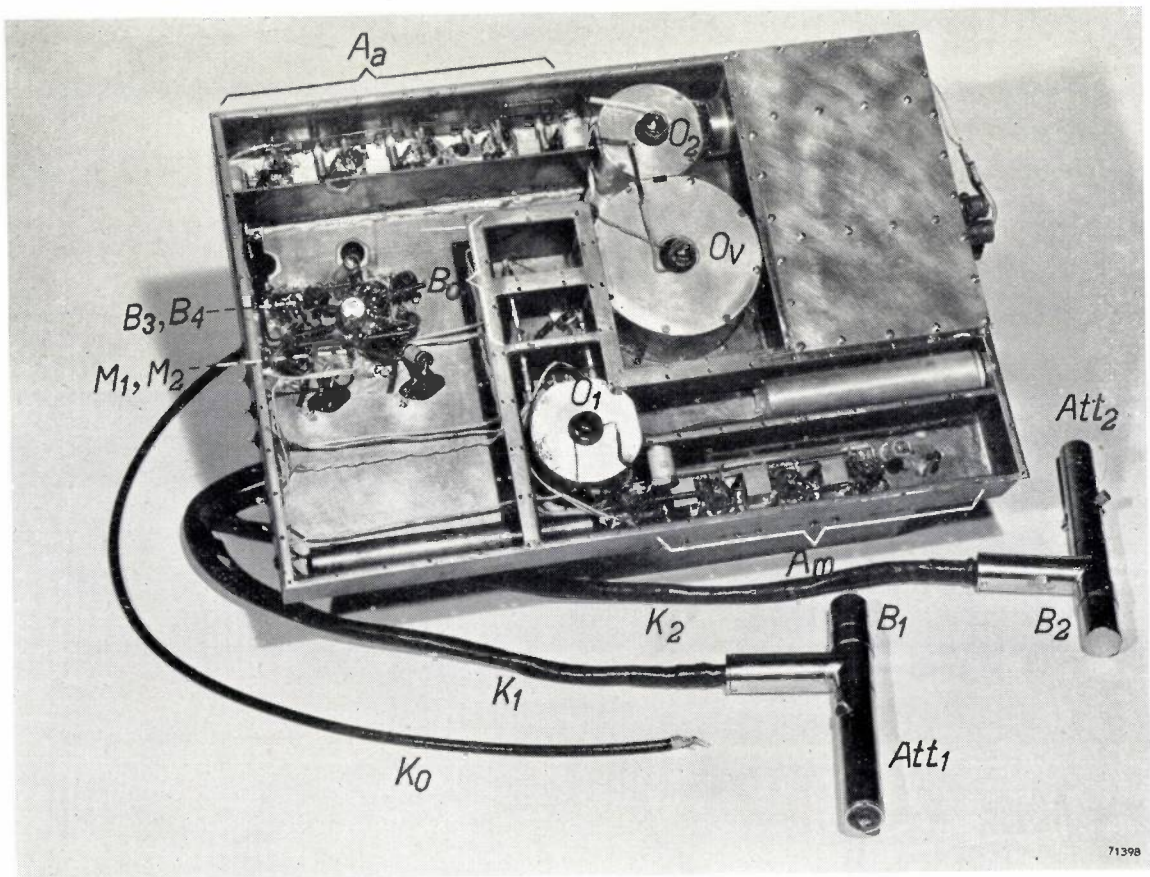
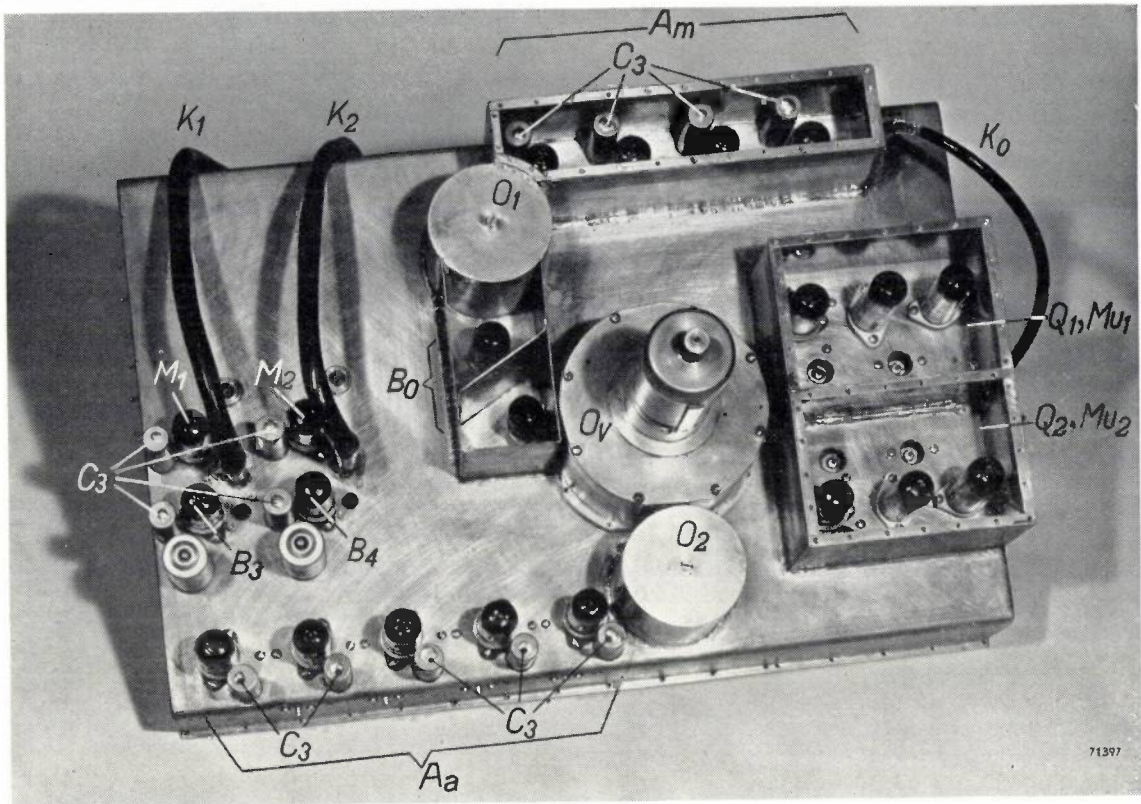


Fig. 5. H.F. chassis (covers taken away). (a) front view, (b) back view. C_3 by-pass capacitors of special type (see fig. 7). Other letters as in fig. 2.

To conclude this general description we give in *fig. 4* a photograph of the complete instrument. A slightly different version, developed at Philips Telecommunication Industries, Hilversum, Netherlands, is described elsewhere ⁶⁾.

Constructional details of the H.F. circuit

The high-frequency chassis is shown in *fig. 5a* and *b*. Various components will be discussed below.

Crystal oscillators and frequency multipliers

The crystal oscillators O_1 and O_2 (*fig. 2*) operate at frequencies of 200.3/27 and 200.0/27 Mc/s, respectively. These frequencies must differ by a fixed amount in spite of temperature fluctuations, by which they are slightly affected. In order to keep this difference as constant as possible, the two quartz crystals have been placed in a common brass block, so that they are subjected to the same variations of temperature. In this way f_i is kept within 299.5 and 300.5 kc/s.

The frequency multipliers, of quite normal design, each consist of three tripler stages. The output valve of each multiplier is inductively coupled with the resonator of its respective oscillator, thereby keeping the latter synchronized.

Combined oscillator-mixers O_1, O_2

The oscillators O_1, O_2 and O_v are all of the same type except that O_v is not synchronized. The resonator (*fig. 6a*), of the so-called foreshortened-coaxial type, has a concentrated capacitance between the outer wall 2 and the flange 6 forming an extension of the inner wall 1. Further details are explained in the subscript of the figure.

⁶⁾ G. Thirup and F. A. Vitha, to appear shortly in Communication News.

In the oscillators O_1 and O_2 , the oscillator valves operate at the same time as mixers. At the inputs, both the self-generated voltage and the voltage from the multiplier are fed in push-pull; on the output side, however, the valves are connected in parallel ⁷⁾. This is shown in *fig. 6a*, but more clearly in *fig. 6b*, where the cavity resonator has been replaced by a circuit with lumped $L-C$ elements. If the system is completely symmetrical, in this way only the frequencies $mf_v \pm n \times 200.3$ or $mf_v \pm n \times 200.0$ Mc/s occur in the output of O_1 and O_2 , respectively, where m and n are confined to *odd* integers. In an asymmetrical circuit, on the contrary, m and n can also assume even values. So by aiming at symmetry, we get rid of many undesirable modulation products. For further details the reader is referred to the Appendix.

200-300 Mc/s buffer stage (B_0)

As shown in *fig. 2*, a buffer stage B_0 has been inserted between the oscillators O_1 and O_v . Its object is to prevent a signal from O_1 reaching O_2 . If, for example, the frequency 200.3 Mc/s could reach O_2 , it would mix with 200.0 Mc/s at that point and produce a spurious intermediate-frequency signal that would interfere with the measurement. In the opposite direction, from O_v to O_1 , the buffer must readily transmit the frequency f_v , ranging from 201.3 to 200.3 Mc/s.

The buffer B_0 consists of two type EC 80 triodes ⁸⁾ in cascade, with grounded grids. The ratio between the transfer

⁷⁾ This is a modification of a mixing system described by A. van Weel, Philips techn. Rev. 8, 193-198, 1946, in particularly *fig. 2*.

⁸⁾ K. Rodenhuis, Two triodes for reception of decimetric waves, Philips techn. Rev. 11, 79-89, 1949.

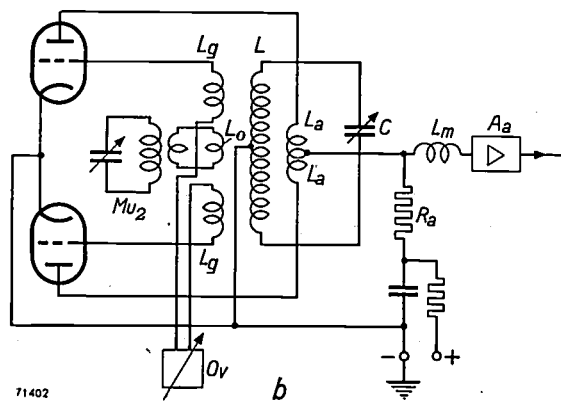
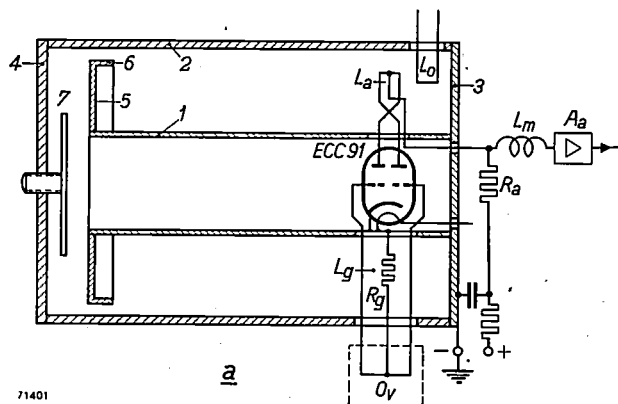


Fig. 6. a) Sectional view of oscillator O_2 (*fig. 2*). Cylinders 1 and 2, bottom 3 and lid 4 form a coaxial cavity resonator. 1 ends in a disc 5 with a rim 6 forming a capacitor with the outer cylinder 2. The movable disc, 7, is provided as a trimmer.

L_g grid circuits, L_a anode circuits of a double triode ECC 91 in push-pull arrangement, coupled with the field in the resonator. L_0 loop fed by the frequency multiplier (Mu_2 ; *fig. 2*), which keeps the frequency at the right value (200.0 Mc/s). The loops L_g have been extended outside the resonator and are coupled with the field of the oscillator O_v with frequency f_v .

The output (frequency $f_a = f_v - 200.0$ Mc/s) is taken from the common anode resistor R_a and supplied, via an inductance L_m (which counteracts the effect of stray capacitances), to the amplifier A_a .

b) shows what the diagram (a) would be if an $L-C$ circuit could be used instead of the resonator.

admittance from cathode to anode to that in the opposite direction is, per stage, approximately $1/\mu$, μ being the amplification factor of the valve. The value of μ for an EC 80 triode is 80. The two valves in cascade yield a ratio of 6400 (76 dB), which has appeared quite sufficient.

The buffer stage can be seen in the middle of figs. 5a and b.

Wide-band amplifiers

The bandwidth to be covered by the amplifiers A_m and A_a (fig. 2) is considerable: it amounts to 100 Mc/s. Fortunately, the requirements with regard to the amplitude and phase characteristics within this band are not stringent, a deviation of ± 3 dB from the flat (low-frequency) part of the characteristic being tolerated, and the phase shift being of no importance at all.

For a conventional one-stage amplifier, with a resistance R in the anode circuit and a stray capacitance C in parallel with R , the frequency at which the gain-frequency characteristic has dropped 3 dB below the flat part is $1/(2\pi RC)$. This frequency limit can be raised by inserting a suitable inductance in the lead from the anode of the valve to the grid of the next valve⁹). This simple measure is made use of at several points of the circuits (for instance, L_m in fig. 6), but in the case of the amplifier it would not yield a bandwidth as wide as required in our case. To meet the demand we have made use of a more complicated circuit, due to Boghosian¹⁰). It is shown in fig. 7. With a

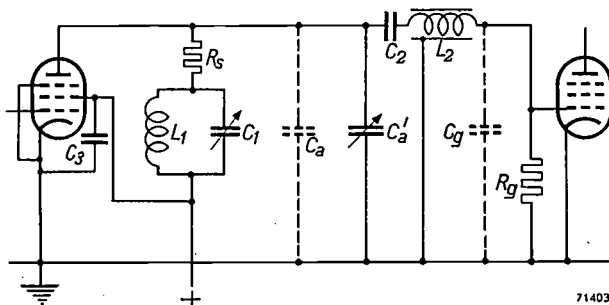


Fig. 7. Circuit diagram (due to Boghosian¹⁰) of one stage of the wide-band amplifiers A_m and A_a (fig. 2). C_a is the anode and wiring capacitance, C_a' the capacitance of a trimmer, C_g the input capacitance and R_g the input resistance of the next valve, C_3 a by-pass capacitor.

For correct functioning, the following relations have to be satisfied:

$$C_a + C_a' = \frac{1.6}{2\pi f_0 R_5}, \quad C_g = \frac{2.0}{2\pi f_0 R_5}, \quad C_1 = \frac{1.067}{2\pi f_0 R_5},$$

$$L_1 = \frac{2R_5}{2\pi f_0}, \quad L_2 = \frac{0.6R_5}{2\pi f_0}.$$

Here f_0 is a frequency 10% higher than the bandwidth.

⁹) See, for instance, H. J. Lindenhovius, G. Arbelet and J. C. van der Breggen, Philips techn. Rev. 11, 206-214, 1950.

¹⁰) Described in chapter 17 of H. W. Bode's book mentioned in footnote 1).

pentode EF 42 (mutual conductance 9 mA/V) and $R = 200 \Omega$, the gain per stage is 1.8, and if the relations between the various circuit parameters mentioned in the subscript of fig. 7 are satisfied, the amplitude characteristic is flat to within a few percent at frequencies up to 100 Mc/s. So five stages in cascade can be used without an excessive reduction of gain at the higher frequencies.

Various circumstances complicate the operation of the circuit. The coil L_2 , for instance, has a self-capacitance, which at very high frequencies makes the reactance higher than that of the pure inductance. This difficulty can be overcome by screening the coil, so that most of the self-capacitance appears in parallel with the capacitances C_a and C_g (see fig. 7). Furthermore, the impedance of the by-pass capacitor C_3 must be low at all frequencies from 1 to 100 Mc/s. A specially wound capacitor, with paper dielectric, was designed for the purpose, in order to avoid resonances in its windings and to keep the inductance low. The capacitor is housed in a brass can, which forms the earthed side and is soldered into a hole in the chassis as near as possible to the screen-grid contact of the valve socket. In this way the inductance of the by-pass connection between this contact and the chassis, via the capacitor, is kept as low as — or even lower than — the inductance of the leads inside the valve. In fig. 5 several of these capacitors are visible.

The amplifier A_m has four stages of the type mentioned and produces an output voltage of maximum 0.1 V across 70 Ω ; A_a , with five stages and a higher input, supplies about 1.5 V to the buffers B_3 and B_4 . In both amplifiers, a circuit tuned to the intermediate frequency f_i has been inserted in one of the cathode leads. Thus heavy negative feedback exists for f_i , resulting in a low gain of unwanted I.F. voltages (see Appendix).

Attenuators At_1 , At_2 and buffers B_1 , B_2

The mixing circuits M_1 and M_2 require input voltages whose levels do not differ too widely. The input and output voltages of the network under test, however, may differ from one another quite considerably in either sense. If, for instance, the network is an amplifier and f_m is within its working frequency range, the output voltage may be some 30 or 40 dB higher than the input voltage. If, on the other hand, f_m is far beyond the working frequency band, the network will in general have a loss of, say, 10 dB. To make up for these differences, a variable attenuator must be introduced between each side of the network under test and the corresponding mixer. These attenuators are variable in five steps of 10 dB each.

The phase shift that these attenuators inevitably introduce should be independent of the attenuator

setting. This can best be accomplished with attenuators of the capacitive type, essentially consisting of a potentiometer formed by two capacitors in series (C_4 and $C_5 + C_6$, *fig. 8*). The attenuation desired can be set up by varying one of the capacitances. In the present case, where a high input impedance is desired, the system with a (small) variable series capacitance is to be preferred.

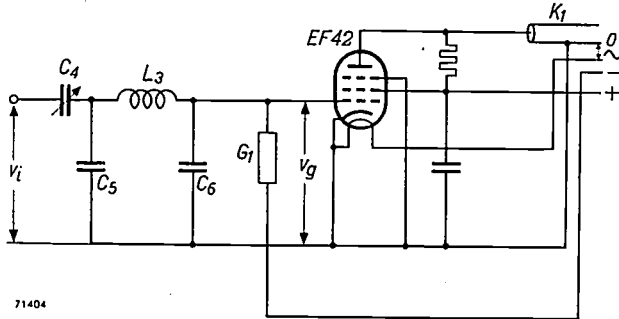


Fig. 8. Diagram of the attenuator Att_1 and the buffer B_1 . C_4 variable capacitor of the sliding type. C_5 wiring capacitance. L_3 wiring inductance. G_1 input conductance (including grid leak) of the buffer valve. C_6 input capacitance of this valve. K_1 cable (140 Ω).

In *fig. 8*, C_4 is the variable series capacitor of the attenuator, C_5 the fixed capacitance of the wiring, C_6 the fixed input capacitance of the subsequent valve (buffer valve), L_3 the inductance of the wiring and of the leads inside the valve, and G_1 the grid leak conductance (including the input conductance of the valve).

At the lower end of the measuring frequency range (f_m near 1 Mc/s), the influence of L_3 — estimated as 0.05 μH — may be neglected. For the ratio between the voltage v_g at the grid of the buffer valve and the input voltage v_i we then find:

$$\frac{v_g}{v_i} = \frac{j\omega_m C_4}{G_1 + j\omega_m (C_4 + C_5 + C_6)}, \dots \quad (1)$$

where $\omega_m = 2\pi f_m$. At these lower frequencies, G_1 may be put equal to the conductance of the grid leak = $10^{-6}\Omega^{-1}$. With $f_m = 1$ Mc/s, $C_5 = 2.5$ pF and $C_6 = 15$ pF, we find from (1) that a variation of C_4 from 2.5 pF downwards to zero causes a variation in phase shift of less than 0.1° —an amount small enough to be of no consequence.

At the upper end of the f_m range the influence of L_3 can no longer be neglected. We then find a more complicated formula. Moreover, G_1 is now dependent on frequency (mainly due to inductance in the cathode lead and to transit time effects). At $f_m = 100$ Mc/s, G_1 is estimated as $10^{-3}\Omega^{-1}$. The phase shift produced by again varying C_4 from 2.5 pF down to zero is then found to be slightly less than 1° , which is still acceptable.

For a capacitor with a maximum capacitance of 2.5 pF and a minimum capacitance low enough to provide the necessary attenuation of 50 dB, the sliding type is to be preferred to the rotary type. It has been given the form of a probe¹¹ (*fig. 9*). The variable capacitor, together with the buffer valve, is housed in an earthed metal tube. One electrode, in the shape of a rod, is fixed and is connected to a protruding pin; the latter makes contact with the input or output terminal of the network under test. The other is cylindrical and can be moved coaxially with respect to the fixed electrode. As explained elsewhere⁹), the attenuation varies almost logarithmically with the distance, provided this distance is large compared with the internal diameter of the tube.

The moving electrode can be located by a detent at five calibrated settings, covering a range of 10-50 dB; the detent consists of a steel ball pressed by a

¹¹) A capacitive attenuator of a similar type, intended for changing the range of a H.F. valve voltmeter, was described at the end of the article mentioned in footnote ⁹).

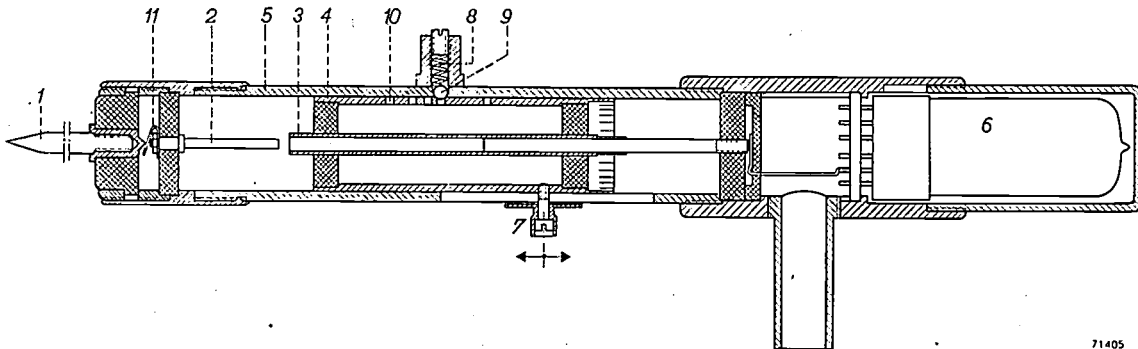


Fig. 9. Sectional view of the probe containing an attenuator and a buffer according to *fig. 8*. 1 contact pin. 2 fixed electrode. 3 tubular electrode mounted in a sliding part 4. 5 earthed metal tube. 6 buffer valve EF 42. 7 knob for moving 4. 8 spring pressing a steel ball 9 into one of the holes 10 to locate the sliding assembly in one of the five calibrated positions. A leaf spring 11 establishes a flexible contact between 1 and 2 to protect the calibration of the attenuator from disturbance by accidental damage to the pin 1.

spring into a series of holes drilled in the moving electrode assembly.

The input capacitance is only 4.5 pF at low and 2.0 pF at high attenuations.

The sliding electrode is connected to the control grid of the buffer valve, a type EF 42 pentode. This valve feeds into a cable, which connects the probe to the instrument itself and which also contains the leads for the heater current and the D.C. supply. The characteristic impedance of the cable is 140 Ω.

Further details may be gathered from fig. 9.

Buffer stages B₃, B₄ and mixing circuits M₁, M₂

The buffer B₄ and the mixer M₂ (fig. 2) together form a unit, the diagram of which is shown in fig. 10. In each circuit a type EF 42 pentode is used. A parallel circuit L₄-C₇, tuned to f_i, is connected in series with the probe cable. Its

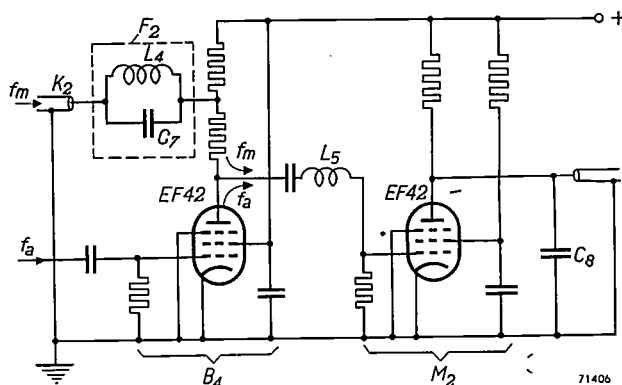


Fig. 10. Diagram of buffer B₄ and mixer M₂. K₂ cable arriving from one of the probes. L₄-C₇ filter (F₂ in fig. 2) suppressing f_i. L₅ inductance for counteracting stray capacitance effects. C₈ input capacitor of low-pass filter F₁ (fig. 3).

function is to suppress the intermediate frequency (see Appendix). An inductance L₅ has been inserted in the lead from the anode of the buffer valve to the control grid of the mixing valve to counteract the adverse effect of stray capacitances⁹⁾.

Stages B₃ and M₁ are similar to B₄ and M₂.

Constructional details of the I.F. circuit

Attenuator Att₃

By means of the attenuator Att₃ (fig. 3) the two intermediate-frequency voltages to be compared

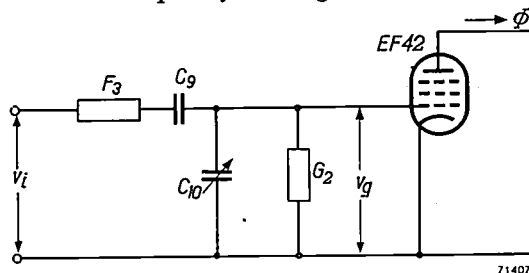


Fig. 11. Attenuator Att₃ (fig. 3) consisting in essence of a fixed capacitor C₉ and a variable capacitor C₁₀. F₃ low-pass filter. G₂ input conductance of a pentode EF 42 feeding the bridge circuit of the phase shifter Φ (see fig. 13a).

can be made equal in amplitude without appreciably changing the phase difference. The attenuator consists of the air capacitor C₁₀ in combination with the other components shown in fig. 11. In contrast to the attenuators Att₁ and Att₂ (fig. 8), the parallel capacitor C₁₀ is variable and the series capacitor C₉ is fixed.

The reason for this choice is the following. The desired maximum attenuation of 20 dB can be obtained either with a parallel capacitor of 500 pF maximum or with a series capacitor in the order of 10 pF. The former is of a standard type with rotary plates, and in the latter case the low capacitance would necessitate a sliding movement — as shown in fig. 9 —, in order to provide sufficient separation of the electrodes. For panel mounting the rotary type is much to be preferred.

In this circuit the frequency is constant (0.3 Mc/s). The filter F₃ (fig. 11) can be represented by a resistance of 900 Ω and the conductance of the grid leak is 10⁻⁶Ω⁻¹. By a proper choice of C₉ and of the extreme values of C₁₀ (40-500 pF), the attenuation can be varied over a range of 20 dB (to be read from the dial of C₁₀), without introducing an error in phase shift of more than 0.5°.

Phase shifter

The principle of the phase-shifting device (Φ in fig. 3) is shown in fig. 12. Two similar loops, l₁ and l₂, are placed perpendicular to each other inside a metal tube. It can be shown by waveguide theory

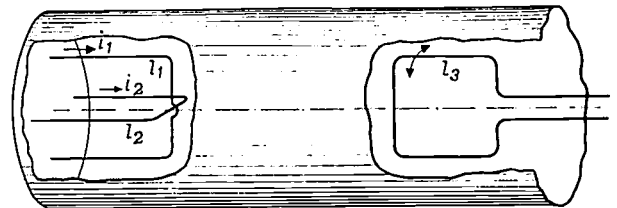


Fig. 12. Phase shifter. Alternating currents i₁ and i₂ of equal amplitudes and in quadrature flow through two similar loops l₁ and l₂ placed perpendicular to each other inside a metal tube. At some distance, a rotating magnetic field of circular symmetry is produced, which acts upon a rotatable loop l₃, placed therein. When l₃ is turned about its axis, the voltage induced in the loop remains constant in amplitude, but is shifted in phase by an amount equal to the angle of rotation, which is readable on a dial.

that if the loops carry alternating currents of equal strengths but in quadrature, then beyond a certain distance from the loops there is a rotating magnetic field with circular symmetry. In this field a third loop, l₃, is placed. In it a voltage is induced, proportional in amplitude to the currents in l₁ and l₂. If l₃ is rotated about the axis of the tube, the induced voltage remains constant in amplitude but varies in phase by an amount equal to the angle over which the loop has been turned.

As we shall see presently, the currents for l_1 and l_2 are supplied by the valve shown in fig. 11. In this way the voltage derived from the input of the network under test can be balanced against that derived from the output of the network. The latter voltage appears at the output of filter F_4 (fig. 3).

It only remains to be explained how the currents for the loops l_1 and l_2 are obtained from the valve EF 42 in fig. 11. The valve feeds a bridge circuit $C_{11}-R_1-C_{12}-R_2$ (fig. 13a), from which are taken

P_2 , they produce in the anode circuits alternating currents that are also in quadrature and — provided the transconductances of the valves are the same — of equal amplitude. The currents supplied by the transformers Tr_1 and Tr_2 , therefore, can be fed to the loops l_1 and l_2 of the phase shifter. As, however, the bridge elements and the intermediate frequency may deviate slightly from the required values, and the transformers and the valves may differ in their respective characteristics, means are necessary to correct for these deviations. Therefore, the grid voltages are not taken from the points B and C (fig. 13a), but from sliding contacts on the resistors R_1 and R_2 . The values of the bridge elements are so chosen that the angle BOC (fig. 13c) is somewhat greater than 90° . By moving the contact on R_1 , the grid voltage of valve P_1 is varied in amplitude but not in phase, and by moving the contact on R_2 the grid voltage of P_2 is varied in phase without an appreciable change of amplitude.

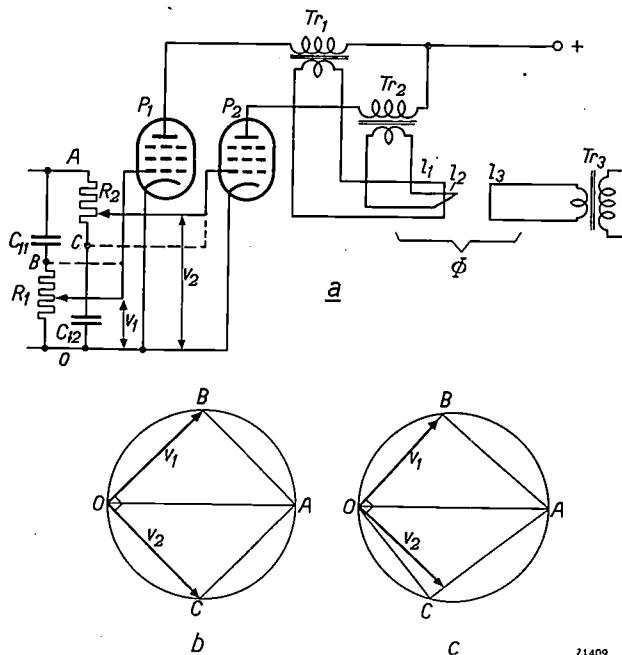


Fig. 13. a) Circuit providing the loop currents for the phase shifter Φ . $R_1-C_{11}-R_2-C_{12}$ bridge fed by the pentode shown in fig. 11. P_1, P_2 two other pentodes EF 42. Tr_1, Tr_2 current transformers. Tr_3 step-up transformer. The transformers Tr_1, Tr_2 and Tr_3 have Ferroxcube cores.

b) If $R_1 = R_2 = 1/(\omega_i C_{11}) = 1/(\omega_i C_{12})$, the vectors representing the voltages across the bridge arms form a square $OBAC$. The voltages OB and OC can then be used as grid voltages, as shown in (a) by dotted lines.

c) In order to allow for trimming, the circuit parameters are so chosen that the angle BOC is slightly greater than 90° , and the grid voltages v_1 and v_2 are taken from sliding contacts on R_1 and R_2 . These contacts can be so adjusted that v_1 and v_2 are equal in amplitude and at right angles to each other.

the grid voltages for a pair of pentodes P_1 and P_2 . Transformers in the anode circuits of P_1 and P_2 match the high impedances of the valves to the low impedance of the loops (each loop consists of one single turn only, to avoid the excitation of undesirable modes in the waveguide).

If the resistances R_1 and R_2 and the capacitances C_{11} and C_{12} of the bridge circuit are such that $R_1 = R_2 = 1/(\omega_i C_{11}) = 1/(\omega_i C_{12})$, then the vectors representing the voltages across the bridge arms form a square ($OBAC$, fig. 13b). The voltages OB and OC are thus in quadrature and equal in amplitude. Used as grid voltages for the valves P_1 and

The circuit can be adjusted in the following way. A voltmeter is connected across the secondary winding of a transformer Tr_3 which steps up the voltage induced in the loop l_3 . A reading of this meter is taken with l_3 turned into the position where it is parallel with l_2 . Next, l_3 is turned parallel with l_1 , and the sliding contact on R_1 adjusted until the same reading of the voltmeter is obtained. If now l_3 is rotated over 360° , the voltage will in general not remain constant during the rotation, due to the loop currents not being exactly in quadrature. In fact, if the voltmeter reading is plotted in polar co-ordinates as a function of the angle of rotation, an ellipse is obtained. The phase difference is now adjusted by means of the sliding contact on R_2 until the ellipse becomes a circle.

With this adjustment, even a slight content of second harmonic may be very troublesome. For this reason, a milliammeter in series with the diode detector in the amplifier A_{IF} can be used with advantage instead of the voltmeter across Tr_3 .

Other components

The voltage induced in l_3 and stepped up by transformer Tr_3 is applied to the control grid of a type EF 42 pentode. The voltage from the filter F_4 (fig. 3) is applied to the control grid of another EF 42 pentode. Since the anodes of these valves are connected in parallel, in the common anode lead an alternating current is obtained which vanishes if the grid voltages are equal in amplitude and opposite in phase, i.e. if the two intermediate-frequency voltages to be compared are made to balance.

The rest of the intermediate-frequency circuit and the audio-frequency circuit (fig. 3) are very similar to the corresponding parts of a normal broadcast receiver, so they need no further description.

Corrections

A condition — so far mentioned only in passing — which has to be fulfilled is that the channels $Att_1-B_1-K_1-F_1-M_1$ and $Att_2-B_2-K_2-F_2-M_2$ (fig. 2) must be equal in transmission properties. The

same applies to the channels B_3-M_1 and B_4-M_2 (fig. 2) and to the channels $M_1-F_3-Att_3-\Phi-A_{IF}$ and $M_2-F_4-A_{IF}$ (fig. 3). For this reason, corresponding components in each of these pairs of channels have been given identical electrical properties as far as possible. Valves, for instance, have been selected for equality in slope, and for F_3 and F_4 low-pass filters (cut-off frequency 0.35 Mc/s) have been preferred to band-pass filters, as the transmission properties of the former are less dependent on temperature and on possible variations of the intermediate frequency.

In spite of all care, however, some asymmetry remains, as can be found by connecting Att_2 to the same point as Att_1 and otherwise using the instrument in the normal way: then, as a rule, the intermediate-frequency voltages do not balance, as they should in case of exact symmetry. These residual amplitude and phase distortions have been measured and plotted as functions of frequency. The curves so obtained show the corrections which have to be applied to the dial readings, if the utmost precision is needed. The corrections never exceed 3% in the amplitude ratio and 3° in the phase angle. The amplitude correction varies somewhat with time owing to ageing of the valves, but by readjusting a trimmer potentiometer the original correction curve can be obtained again without any difficulty. The correction for phase difference is practically constant with time.

Appendix: Suppression of stray effects

The instrument described can only be expected to function well if the utmost care is taken to avoid stray effects that would falsify the measurements. These stray effects are partly due to the presence of unwanted couplings between the input and output of the network under test, partly also to unwanted frequencies arising at certain points of the circuit. As will appear from the survey given below, the remedies consist in screening and in the insertion of various buffers and filters.

Necessity of buffers B_1 , B_2 , B_3 and B_4

In the absence of the buffers B_1 , B_2 , B_3 and B_4 (fig. 2), a coupling by way of Att_1 , F_1 , F_2 and Att_2 would exist between the input and output of the network under test. Therefore, the measurements would not be performed on the original network but on this network plus the extra coupling. As a means of suppressing this coupling in the direction from input to output, the buffers B_3 and B_2 have been inserted; they are networks transmitting readily

in the direction of the arrows but presenting a high attenuation the other way. Similar buffers, B_4 and B_1 , prevent a coupling from output to input.

A further purpose of the buffers B_1 and B_2 is to prevent a signal with the auxiliary frequency f_a , which is present at the input terminals of the mixers M_1 and M_2 , from reaching the network under test. Here, owing to its large amplitude (see below), this signal would make the valves operate over a non-linear part of their characteristics and so give rise to unwanted modulation products.

Necessity of buffer B_0 and of screening between the generators

If a signal with the frequency 200.0 Mc/s, present in the generator G_2 , (fig. 2), were to reach the oscillator-mixer O_1 , then, by mixing with signals of frequencies 200.3 Mc/s and f_v , it would cause the formation of various unwanted frequencies. Similarly, a signal with the frequency 200.3 Mc/s reaching O_2 from G_1 would be the cause of unwanted frequencies being produced in O_2 . In the table below, the wanted frequencies and some unwanted ones are mentioned. All these modulation products are amplified by the amplifiers A_m and A_a . The 0.3 Mc/s component passes directly on to the I.F. circuit and causes a wrong measuring result. Moreover, in the mixers M_1 and M_2 the unwanted components with the frequencies f_m and f_a are added to the wanted components, which also disturbs the measurement.

Table, showing some of the frequencies produced in the oscillators-mixers O_1 and O_2 , the unwanted ones by interaction between O_1 and O_2 .

	Output frequencies from O_1 Mc/s	Output frequencies from O_2 Mc/s
Wanted	$f_v - 200.3 = f_m$	$f_v - 200.0 = f_a$
Unwanted "first order" modulation products	$200.3 - 200.0 = f_i$	$200.3 - 200.0 = f_i$
Unwanted "higher order" modulation products of importance	$f_v - 200.0 = f_a$	$f_v - 200.3 = f_m$

The formation of unwanted output components in O_1 and O_2 can be suppressed at the root of the evil by careful screening, combined with the insertion of a buffer (B_0) between O_1 and O_2 , and further by the insertion of 0.3 Mc/s band-stop filters in the amplifiers A_m and A_a .

Signal levels

Fig. 14 gives an idea about the necessary differences in signal level at various points of the circuit. We start from the following two requirements:

- 1) At the output terminals of M_1 and M_2 a difference of at least 50 dB between the wanted and unwanted I.F. components is necessary in order to keep the modulus and phase errors, due to the presence of unwanted I.F. components, sufficiently small. For M_2 this is shown in fig. 14, below point a .

this latter mutual conductance is some 10 dB higher than the conversion transconductance. For this reason, at the input of M_2 (b , fig. 14), the parasitic I.F. component must be at least $50 + 10 = 60$ dB in level below the f_m component, i.e. 100 dB below the f_a component.

Let us now consider the amplifier under test. As the filter F_2 provides an attenuation of 60 dB at the frequency f_i , it is admissible that at the output terminals of the amplifier (c , fig. 14) the (unwanted) f_i component has the same level as the (wanted)

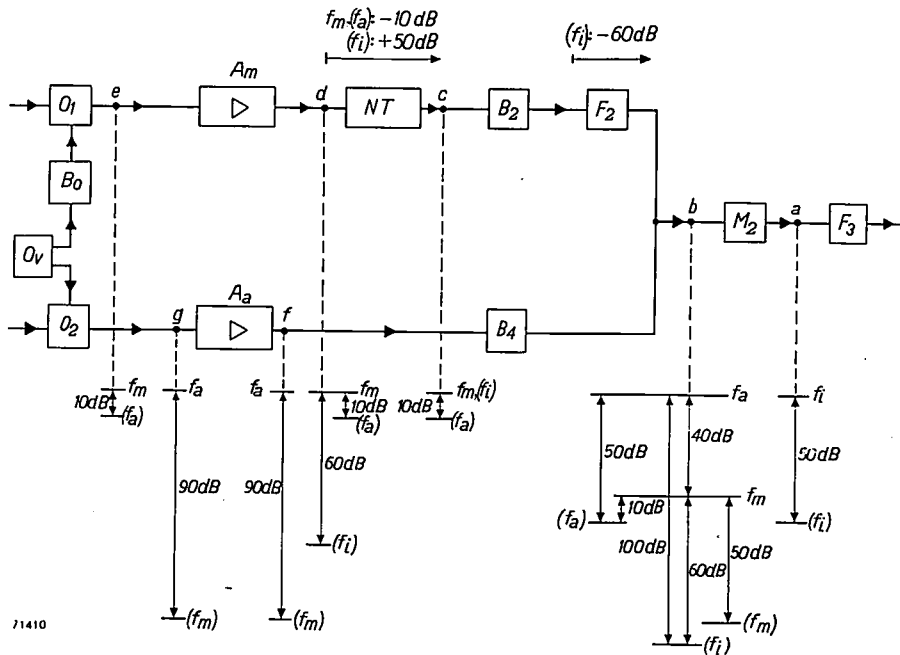


Fig. 14. Part of the block schematic fig. 2 redrawn, showing levels of components of wanted and unwanted frequencies at points indicated by dotted lines. Symbols of unwanted frequencies are placed in brackets.

- 2) For a proper linear working of M_1 and M_2 , at the input terminals of these mixers, the f_a component level should be 40 dB higher than the f_m component level. This, too, is shown for M_2 in fig. 14, below point b .

From these requirements it can be derived what differences in level are necessary at various points of the circuit.

Levels of unwanted I.F. components

The wanted I.F. output components of M_1 and M_2 are due to the mixing of an f_m and an f_a input component, so they are proportional to the conversion transconductance of the mixing valve. The unwanted I.F. components, however, are due to direct amplification of parasitic I.F. input components, and so they are proportional to the mutual conductance of the mixing valve. Now

f_m component. At the input terminals (point d), however, a difference of 60 dB is necessary, as the amplifier under test may provide a gain of 50 dB for signals with frequency f_i while attenuating some 10 dB those with frequency f_m .

The difference of 60 dB at d between the f_m component and the parasitic f_i component, and the difference of 100 dB at c between the f_a component and the parasitic f_i component are obtained by the measures already stated: partly by reducing the gain of the amplifiers A_m and A_a for I.F. signals (by means of selective feedback), and partly by preventing interaction between the oscillators O_1 and O_2 .

Levels of unwanted H.F. components

We now draw our attention to the presence of unwanted high-frequency components in the outputs of the oscillators, i.e. the presence of f_a at O_1 , and

the presence of f_m at O_2 . As these frequencies lie very close together, the response of the amplifiers, buffers and filters to a signal of frequency f_a will be substantially the same as that to a signal of frequency f_m .

Returning now to point *b* in fig. 14, we have a signal of frequency f_a arriving legally by way of A_a and B_a , and an unwanted signal of the same frequency arriving via A_m , NT , B_2 and F_2 . The phase difference between these signals can have any value and will depend on f_a and on the properties of the network under test. In order to keep the resulting f_a signal at *b* sufficiently constant in phase under all circumstances, the level of the parasitic f_a component must be at least 50 dB below the level of the legal f_a component, i.e. 10 dB below the level of the legal f_m component (see fig. 14). The same holds for points *c*, *d* and *e*. This requirement is not difficult to meet.

For a similar reason, at *b* the parasitic f_m signal must be 50 dB below the legal f_m signal, i.e. 90 dB below the f_a signal, and this holds also for points *f* and *g*. This difference of 90 dB is a rather severe requirement. It is mainly met by blocking the way from O_1 , to O_2 via O_v (attenuation of 76 dB by buffer B_0).

Higher-order modulation products of minor importance

Neglecting the presence of parasitic signals at the input terminals of O_1 and O_2 , we can write the frequencies at the output terminals as

$$mf_v \pm n \times 200.3 \text{ Mc/s} \quad \dots \dots \dots (2)$$

and

$$pf_v \pm q \times 200.0 \text{ Mc/s}, \quad \dots \dots \dots (3)$$

respectively. In the general case the coefficients m , n , p , and q are 0, 1, 2, 3,... Frequencies corresponding to the + signs in (2) and (3) can be left out of consideration, as they are well above the frequency range of the amplifiers A_m and A_a .

Mixing of the series (2) and (3) in M_1 and M_2 produces at the output terminals of the HF circuit a new series of frequencies:

$$s(pf_v - q \times 200.0) \pm r(mf_v - n \times 200.3) \text{ Mc/s}, \quad (4)$$

where the coefficients r and s are 0, 1, 2, 3,... In the special case of $m = n = p = q = r = s = 1$, and with the minus sign before r , we find the wanted intermediate frequency

$f_i = 0.3 \text{ Mc/s}$. The other components are only troublesome if their frequencies, given by (4), happen to be in the neighbourhood of f_i and if, moreover, their amplitudes are fairly large. Thanks to a suitable choice of the fixed frequencies (200.0 and 200.3 Mc/s) and of the variable frequency f_v , parasitic intermediate frequencies occur only for rather high values of the coefficients in (4); in other words: only certain modulation products of high orders have frequencies that can be troublesome. Now a high order means a small amplitude, so the trouble experienced is not very great.

Moreover, a substantial reduction in interference is obtained by arranging each of the oscillators O_1 and O_2 in push-pull at the input and in parallel at the output side (see fig. 6). In this way m , n , p and q are limited to odd numbers.

A further help is the aural observation of balance, the ear being able to distinguish the wanted beat note from parasitic ones.

The result is that over the whole range of measuring frequencies only at a few isolated values of f_m some interference is experienced. The unwanted signals considered here do not influence the measuring result, in contradistinction to those discussed in preceding paragraphs.

Summary. Measurement of complex voltage ratios is necessary to obtain the Nyquist curve of feedback amplifiers or other networks. In order to investigate the stability of the amplifier, the frequency range of the measurements must be extended far beyond the highest frequency for which the amplifier is intended. In the present instrument, this range is from 1 to 100 Mc/s.

Two voltages with a fixed intermediate frequency $f_i = 0.3 \text{ Mc/s}$ are derived from the input and output voltages of the network under test by means of a system of frequency conversion. The I.F. voltages have the same complex ratio as the H.F. input and output voltages, whose frequency is, of course, the variable measuring frequency. In a balancing system, one of the I.F. voltages is changed in amplitude and phase — by amounts readable from dials — until it cancels the other. Balance is indicated by the vanishing of a beat note from a loudspeaker. The accuracy is better than 0.2 dB in modulus and better than 2° in phase angle.

The variable measuring frequency f_m is obtained as the difference between a frequency f_v variable from 201.3 to 300.3 Mc/s and the frequency 200.3 Mc/s derived from a crystal oscillator. In a similar way, an auxiliary frequency f_a is obtained as the difference between f_v and the frequency 200.0 Mc/s derived from a second crystal oscillator. So by merely varying f_v , the frequencies f_m and f_a are varied simultaneously, while the difference $f_a - f_m$ is kept constant at 0.3 Mc/s. Voltages with frequency f_a are mixed with the input and output voltages (frequency f_m) of the network under test for producing the I.F. voltages to be compared.

Of the various components described, special mention should be made of the phase balancing device, from the dial of which the phase angle to be determined can be read.

An Appendix deals with the suppression of stray effects that would falsify the measurements.

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.

1993: J. J. Zaalberg van Zelst: Over versterkers met weinig veranderende versterking (T. Ned. Radiogenootschap 16, 117-135, 1951, No. 3). (On amplifiers with almost constant gain; in Dutch.)

This article contains a survey of methods providing a reduction of the variation in gain of amplifiers. The methods using at least some constant network elements are divided into two groups. In those of the first group the influence of the actual change of the constants of the valves is eliminated by corrections of the input or output voltages, in those of the second group the actual change is reduced by regulating voltages derived from an auxiliary signal.

It is also shown to be possible to construct an amplifier having small variations in gain when elements are used which are not constant but have a limited variation. The fundamental limits of this procedure are discussed, regarding oscillators to be selective amplifiers of noise with positive feedback. See Philips techn. Rev. 9, 25-32 and 309-315, 1947.

1994: J. A. Kok: On electrostatic plasma oscillations in metals (Physica 17, 543-547, 1951, No. 5).

In this article an attempt is made to apply to metals the equations of the electrostatic plasma oscillations and of ions in electrical discharges in gases. It is pointed out that for electrons in metals the Fermi-Dirac distribution law has to be used and not the Maxwell statistics. The frequencies of the electrostatic ion oscillations have an upper limit of the same order of magnitude as the Debye maximum frequency.

R 165: H. B. G. Casimir: On the theory of electromagnetic waves in resonant cavities (Philips Res. Rep. 6, 162-182, 1951, No. 3).

In these lectures (Madrid 1949) a survey is given of the theory of standing electromagnetic waves in resonant cavities. The formal analogy between the modes of vibration of a cavity, the modes of vibration of a network of discrete elements, and the vibration of a simple LC circuit is emphasized. Special attention is given to the theory of perturbations; this theory is then applied to a number of examples. These include the determination of the

high-frequency properties of magnetic materials by means of cavities into which small spheres of the material are introduced, and the coupling of two identical cavities by a small hole in a dividing wall. In the last section some remarks are made about the zero-point energy of empty space.

R 166: J. L. Meijering: Segregation in regular ternary solutions, II (Philips Res. Rep. 6, 183-210, 1951, No. 3).

A straightforward method for numerical calculation of tie-lines is given and applied for several combinations of the parameters. When a tie-line has been calculated, the equation furnishes an infinite set of parameter combinations where the same tie-line appears. The initial directions of the binodal curves are given; the widening or narrowing of a binary miscibility gap by a small addition is governed by the difference in affinity of the added substance for the components of the binary system. In certain systems there is a pseudo-binary section. When a , b and c , the coefficients occurring in the thermodynamic potential, are positive, three-phase equilibria appear and the stability of the computed phase diagram must then be checked by means of the spinodal curve. A four-phase transformation of the eutectic type appears when a , b and c are about equally positive. Solubility curves of a pure solid component and their interference with a miscibility gap are also treated. Some remarks are made on multicomponent systems.

R 167: G. Diemer: Microwave diode conductance in the exponential region of the characteristic (Philips Res. Rep. 6, 211-223, 1951, No. 3).

A transit-time theory is expounded for the conductance of a diode in the exponential region of its characteristic. A linear retarding field is assumed, and mutual electron interaction is ignored. For diodes with short distance between cathode and anode the value of this conductance in the voltage range $-2V < V_a < -1V$ may considerably exceed that of the total-emission conductance even in the microwave band. Application of the theoretical results to earlier measurements on "short-distance" diodes disposes of part of the former discrepancy between theory and experiment.

R 168: J. B. de Boer: Visibility of approach and runway lights (Philips Res. Rep. 6, 224-239 1951, No. 3).

In this paper the physiological data governing the visibility of steady lights are surveyed. For circular sources of light extensive information can be obtained from the results of observations published by several investigators. The influence of the shape of the sources and that of neighbouring light sources has been empirically studied for rectangular sources and for sources arranged in rows. Further, a brief account is given of what is known regarding the influence of the colour of the light. All these influences can be allowed for by applying correction factors. To this end, "size", "shape", "row" and "colour" factors have been introduced. The threshold values found in the laboratory have to be multiplied by a safety factor in order to get data suitable for calculating the luminous intensities required in order that a pilot in actual flight will be able to see the lights.

R 169: J. B. de Boer: Calculations of the light distribution of approach and runway lights (Philips Res. Rep. 6, 241-250, 1951, No. 4).

The relation between daylight-visibility in a foggy atmosphere and the transmission of light through fog is treated briefly and qualitatively in so far as data are needed for calculating the required luminous intensity of approach and runway lights. Fundamental physiological information on the visibility of lights through fog, required for the same calculation, has been taken from an earlier paper in these Reports. A method of calculation is described which allows of the luminous intensity of approach and runway lights and its spatial distribution being determined in a simple way (by means of two graphs).

R 170: G. J. Fortuin: Visual power and visibility (Philips Res. Rep. 6, 251-287, 1951, No. 4).

Since several decennia the visual acuity (the reciprocal value of the angular size of the smallest perceptible object) has been the most usual standard of visual capacities, although it was known that its value depends on the examining conditions (the contrast of the test objects, the exposure time, the field brightness, etc.). With due alterations of details the same applies to other standards of visual capacities, such as contrast sensitivity and speed of vision. Nevertheless, the capacity of obtaining information about the electromagnetic behaviour of one's sur-

roundings is a property exclusive to the observer and, therefore, independent of external conditions. In this paper a quantitative definition is given of this capacity, which is called the "visual power". This definition is based on 9120 determinations, so that generally the visual power is indeed independent of the experimental conditions. Subjects, 228 in total, were school children, applicants for employment and employees.

The visual power decreases with increasing age and this decline is already present at ages under twenty years. A further consideration of the results revealed the occurrence of some different visual types, one of them being due to the presence of opacities in the optical media of the eye.

R 171: K. S. Knol: Determination of the electron temperature in gas discharges by noise measurements (Philips Res. Rep. 6, 288-302, 1951, No. 4).

The use of an electric discharge in a gas as a microwave noise source has been studied. It appears that the noise power available is determined by the electron temperature of the discharge. Certain precautions have to be taken in order to obtain a relatively loss-free matching of the discharge conductance to the waveguide. The electron temperatures of He, A and X are measured as functions of the gas pressure and the direct current. A comparison of these measurements is made with a theoretical formula for the electron temperature as a function of the product of the gas pressure and the radius of the discharge tube. The agreement between theory and experiment is satisfactory. The values of the electron temperature determined in the same discharge tube by noise measurements and by probe measurements are equal within the limits of accuracy. From semi-theoretical arguments a relation is derived between the electron temperature in a gas discharge and the longest resonant wavelength of that gas.

R 172: J. M. Luttinger *): Wave propagation in one-dimensional structures (Philips Res. Rep. 6, 303-310, 1951, No. 4).

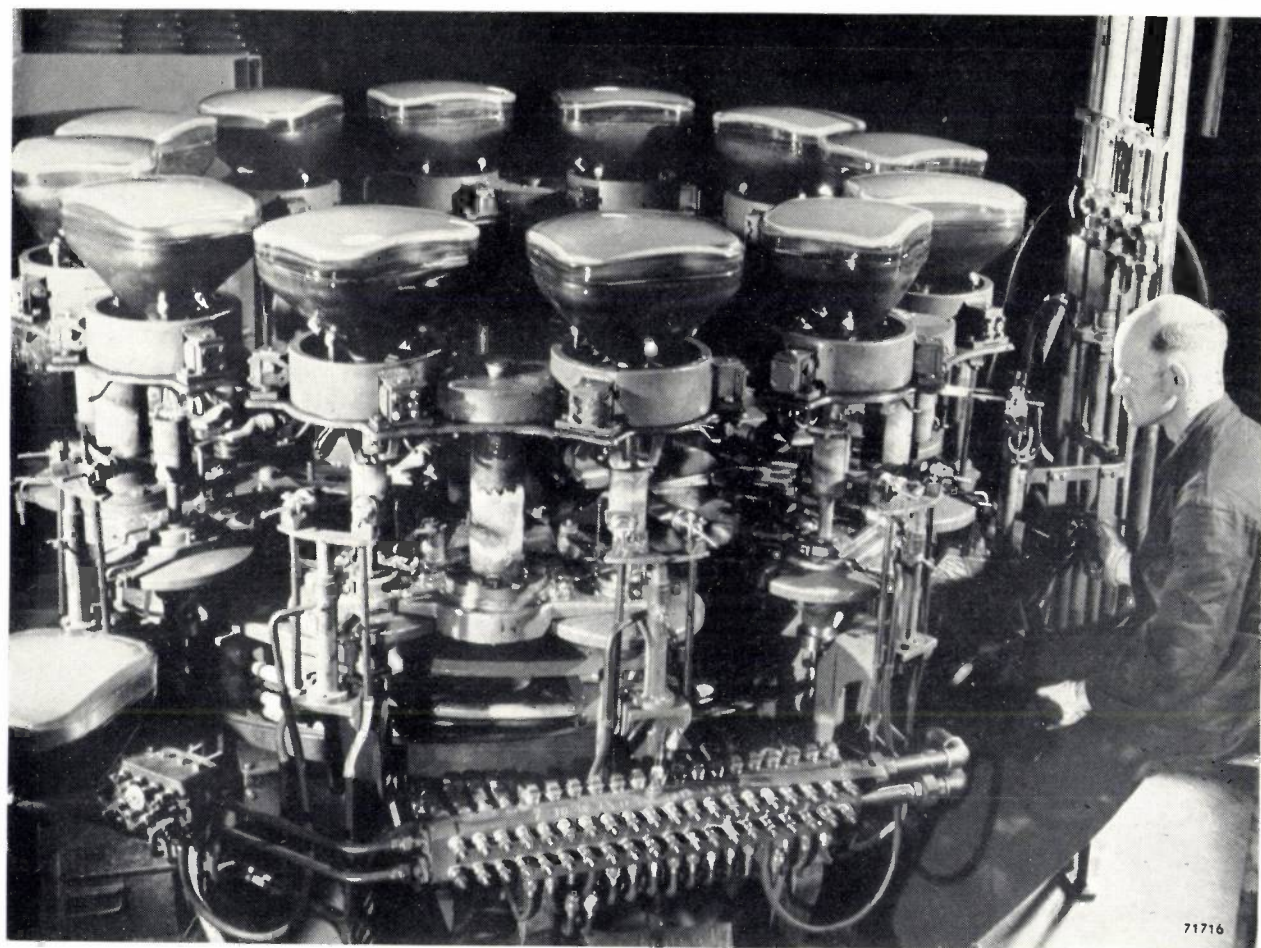
A conjecture of Saxon and Hutner (see No. R 102 of these abstracts) on the forbidden energy levels of any substitutional alloy of a certain one-dimensional crystal model is proved. Analogous results are derived for wave transmission down a line loaded with two-terminals.

*) Dept. of Physics, Univ. of Wisconsin, Madison, Wis., U.S.A.

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



Photograph Walter Nürnberg

ELECTRONIC TUBES

A synopsis by J. L. H. JONKER.

621.385

The following article covers the main points of an address by Prof. Dr. J. L. H. Jonker on his formal acceptance of the office of professor at the Technical University at Delft on 19th March 1952. Although the extent of his subject did not admit of more than a brief review, this "bird's-eye view", which brings many widely separated aspects into a common focus, has an appeal of its own.

Nothing but the greatest admiration can be felt by anyone tracing the rapid development of electronics, that is, electronic tubes and their applications, which has sprung from the phenomenal growth of radio in the first half of the twentieth century.

It was roughly at the turn of the century that Hertz, Sir Oliver Lodge and Marconi carried out the experiments that led to the successful introduction of telegraphy without wires by means of high-frequency oscillations.

In the short span of a few decades radio has

undergone almost unbelievable technical development and improvement, and it has achieved a revolution in the interchange of ideas between individuals and peoples alike. It has come to assume such importance in the lives of almost everyone that throughout the whole evolution of mankind there have been few technical developments to equal it.

The early days of radio

This development of radio, which took the world by storm, founds its origin in the experiments already mentioned. Channels were established across the oceans, first telegraphically and later by telephone, up to the point where radio contact now covers the whole globe.

In the early years, technical knowledge and literature on the subject were in the nature of things very limited. Because the individual experimenter was, however, in very many instances able to achieve valuable results and improvements using only simple and often home-made equipment, many felt themselves drawn towards this sphere and this is undoubtedly one of the causes of the very rapid advances made. Those who were privileged to witness at close quarters some part of this inspiring early period may look back on those exciting days with some regret, comparing them with conditions as they are today, now that improvement and research into specific problems are in the hands of a corps of specialists having the most ingenious laboratory equipment at their disposal.

As a result of the work of these specialists, literature on the subject of radio would now constitute a whole library in itself, whereas, on the other hand, the ceaseless flow of publications presents to the expert the problem of keeping in touch with everything that is of interest in his work. Here, too, a too copious emission results in so great a space charge that the ultimate object is defeated. In the United States this urgent problem has recently given rise to an investigation into the statistical distribution of such publications among the various periodicals; the result of this investigation has been reported — in another publication! — in the periodical containing the largest number of such publications¹⁾.

It is because of the initial circumstances already outlined that history can point to one specific invention of one man, as being as it were the lever which released radio from its trammels. This was the start of the subsequent fantastic growth and was, in the words of the Nobel-prize winner Rabi:

“so outstanding in its consequences that it almost ranks with the greatest inventions of all time”.

This invention is the *triode* of Lee de Forest.

The advent of the triode

In seeking for a better detector of radio signals Lee de Forest²⁾ in 1907 introduced a grid between the cathode and anode of the existing diode devised by Sir John Ambrose Fleming, thus obtaining a means of controlling the flow of electrons between

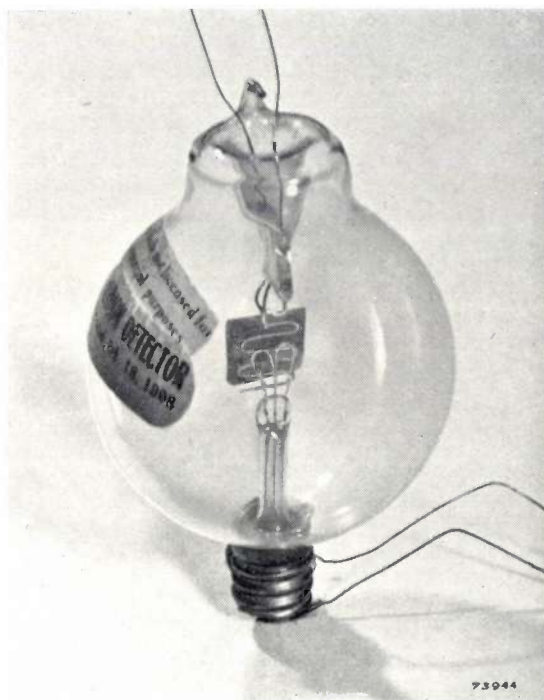


Fig. 1. The original triode (“audion”) of Lee de Forest.

cathode and anode without the consumption of any energy (figs 1 and 2). Years elapsed before the operation and possibilities of this new invention were fully appreciated, in consequence of which De Forest became so short of funds that he had to let his European patents lapse! In 1911 Robert von Lieben gave a demonstration of the amplification properties of the triode at the Berliner Physikalische Gesellschaft. In 1913 different research workers simultaneously invented the “feed-back”, which permitted the triode to be used as a very effective generator of high-frequency oscillations. The last-mentioned invention gave rise to the longest case of patents litigation in this field in history³⁾, a contest which was not settled until 1934, in favour of De Forest.

The invention of the triode, which marked the commencement of the electronic era, gave to radio

at once a better detector, the long sought amplifier and an oscillator. These various possibilities provided the stimulus for a careful study of these tubes and for such modifications as would ensure optimum results in any function in transmitting and receiving equipment, and this has in turn produced the innumerable varieties of electronic tubes which we know today.

reproduction and the oscilloscope. In the years that followed, research penetrated the realms of ultra-short waves, atmospheric interference, frequency modulation and television, apart from innumerable industrial applications of electronics. In each case the electronic tube was the key that opened the door to the new sphere, and often, in new forms, to entirely new possibilities. This

UNITED STATES PATENT OFFICE.

LEE DE FOREST, OF NEW YORK, N. Y., ASSIGNOR, BY MESNE ASSIGNMENTS, TO DE FO. EST RADIO TELEPHONE CO., A CORPORATION OF NEW YORK.

SPACE TELEGRAPHY.

No. 2,792,822.

Specification of Letters Patent.

Patented Feb. 19, 1958

Application filed January 29, 1937. Serial No. 234,662.

To all whom it may concern:

Be it known that LEE DE FOREST, a citizen of the United States, and a resident of New York, in the county of New York and State of New York, have invented a new and useful improvement in Space Telegraphy, of which the following is a specification.

My invention relates to wireless telegraph receivers or oscillation detectors of a type heretofore described in my prior Letters Patent Nos. 824,637, June 26, 1906 and 816,070, November 13, 1906.

The objects of my invention are to increase the sensitiveness of oscillation detectors comprising in their construction a gaseous medium by means of the structural features and circuit arrangements which are hereinafter more fully described.

My invention will be described with reference to the drawings which accompany and form a part of the present specification, although it is to be understood that many modifications may be made in the apparatus and systems herein described without departing from the principles of my invention.

In the drawings, Figure 1 represents in

end brought out to the terminal 3. Interposed between the members F and G is a grid-shaped member A, which may be formed of platinum wire, and which has one end brought out to the terminal 1. The local receiving circuit, which includes the battery B, or other suitable source of electromotive force, and the signal indicating device T, which may be a telephone receiver, has its terminals connected to the plate G and filament F at the points 3 and 4 respectively. The means for conveying the oscillations to be detected to the oscillation-detector, are the conductors which connect the filament F and grid A to the tuned receiving circuit and, as shown, said conductors pass from the terminals 2 and 1 to the armatures of the condenser C.

I have determined experimentally that the presence of the conducting member A, which as before stated may be grid-shaped, increases the sensitiveness of the oscillation detector and, inasmuch as the explanation of this phenomenon is exceedingly complex and at best would be merely tentative, I do not deem it necessary herein to enter into a de-

when said condenser is present over the sounds produced therein under the same conditions when said condenser is not employed.

It will be understood that the circuit arrangements herein described with reference to the particular forms of audion herein disclosed may with advantage also be employed with various other types of audion.

I claim:

1. An oscillation detector comprising an evacuated vessel, an electrode inclosed therein, means for heating said electrode, a second electrode inclosed within said vessel, a local circuit having its terminals electrically connected to said electrodes, a conducting member inclosed within said vessel and located between said electrodes, and means for conveying the oscillations to be detected to the first mentioned electrode and said conducting member.
2. An oscillation detector comprising an evacuated vessel, two electrodes inclosed within said vessel, means for heating one of said electrodes, and a conducting member inclosed within said vessel and interposed between said electrodes.
3. An oscillation detector comprising an evacuated vessel, two electrodes inclosed within said vessel, means for heating one of said electrodes, and a grid-shaped member of conducting material inclosed within said vessel and interposed between said electrodes.
4. An oscillation detector comprising an

ducting member from charged.

7. An oscillation detector comprising an evacuated vessel including gaseous medium members inclosed therein circuit, a circuit of said oscillation circuit members, a condenser inclosed therein, and means for conveying the oscillations to be detected to the first mentioned electrode and said conducting member.

8. An oscillation detector comprising an evacuated vessel, two electrodes inclosed therein, means for heating one of said electrodes, a conducting member inclosed within said vessel and interposed between said electrodes, and means for conveying the oscillations to be detected to the first mentioned electrode and said conducting member.

9. An oscillation detector comprising an evacuated vessel, two electrodes inclosed within said vessel, means for heating one of said electrodes, and a grid-shaped member of conducting material inclosed within said vessel and interposed between said electrodes.

10. An oscillation detector comprising an evacuated vessel including gaseous medium members inclosed therein circuit, a circuit of said oscillation circuit members, a condenser inclosed therein, and means for conveying the oscillations to be detected to the first mentioned electrode and said conducting member.

Fig. 2. Extracts from the United States patent granted to Lee de Forest, covering the invention of his triode. The "grid-shaped member" is mentioned in the third claim.

Rapid development

During and after the first world war we have seen the efforts that were directed towards producing electronic tubes of high and very high power for the establishment of world-wide radio communication. Then came a remarkable offshoot in the field of electronics: the radio broadcast, commenced with noble idealism, but later to be misused for commercial advertising and political propaganda. Radio broadcasting has in many parts of the world called into existence enormous electronic industries whose annual output is now estimated at some 500 million tubes and a good 30 million receivers, the turnover for 1952 being estimated at many gigadollars. In this way a new sphere of employment has been created for hundreds of thousands in the manufacture and application of electronic tubes.

The young but rapidly expanding industries with their many specialists soon found more scope in other directions, to mention only sound amplification systems, soundfilm, improved gramophone

applies equally to the latest branches of electronics, developed during the last war and subsequently perfected, e.g. radar, sonar and other aids to navigation, radio relay stations and computing equipment.

Electronics in general is much indebted to the second world war, that black page in the history of mankind, since technical development was urged forward by the exigencies of the war without being hampered by economic and commercial factors. Many of the results obtained in this way have since proved to be of great scientific, technical and commercial interest and have been adopted or further developed for peaceful occupations; moreover, new techniques are continually being added to the already wide field of electronics, such being for example radio astronomy, colour television and microwave spectroscopy. The very extensive scope of application of electronic tubes and the consequent diversity of their form may be illustrated with reference to some of the extremes among the dimensions involved.

The diversity of present-day electronic tubes

It is possible by means of electrometer triodes to measure currents of 10^{-15} amperes, whilst in transmitting tubes and gas-filled rectifiers, current peaks of 100 amperes or more occur. Using electronic tubes we can measure alternating voltages of 10^{-6} volts (it should be recognized here that in that case out of a possible 10^{18} available electrons in an amplifying tube, only about 10^{10} are controlled). On the other hand, an installation is being built by the Stanford University in California (a "linear accelerator") which, with the aid of enormous transmitting tubes (klystrons), operated at 300 000 V; will accelerate electrons to the equivalent of a thousand million volts⁴). Electronic tubes can be made to operate within a range of from 0 to 10^{11} cycles per second. The weights of conventional types of tube may differ one from the other by a factor of as much as 100 000.

Within these extreme limits there is an enormous variety of design among electronic tubes; a cautious estimate places the number of different commercial types at about 20 000.

It is not within our scope to give even the briefest review of all these different designs. We shall confine ourselves to a short discussion of some of the more important applications only.

Receiving tubes

The most important field of application of tubes numerically is without a doubt to be found in radio receivers. In the last decades there has been a refinement in the design of the

receiving tube, consisting in the attainment of smaller electrodes and closer spacing between the electrodes, that have resulted in an over-all reduction in the size of the tube (*fig. 3*). The inherent advantages are, amongst others, lower current consumption and better short-wave characteristics; manufacturing costs have dropped owing to more rapid production, lower consumption of materials, smaller stores and reduced transport costs. The receivers in which the tubes are used could be made more compact in consequence.

Intensive research has taken place into the many physical phenomena occurring in electronic tubes that may be useful or detrimental, such as thermal, secondary and photo-emission (*fig. 4*), fluorescence, gas discharge, space charge, potential fields, electron paths, noise and, in short-wave operation, transit-time effects and damping. Investigations into these effects have not by any means ended; on the contrary, efforts are still being made, based on a growing knowledge of these often highly complex phenomena, to improve the performance of the tubes. It was, for example, possible to improve the efficiency of pentodes considerably by taking into account in the design the paths of the electrons through the potential fields of the grids⁵). In most cases, however, a computation of these potential fields and the paths of the electrons within them would take up far too much time, or would prove to be impossible or possible only under very restricted conditions. By means of models of the electrode system in the "electrolytic tank" or on a rubber sheet (*figs 5 and 6*), a better idea is obtained

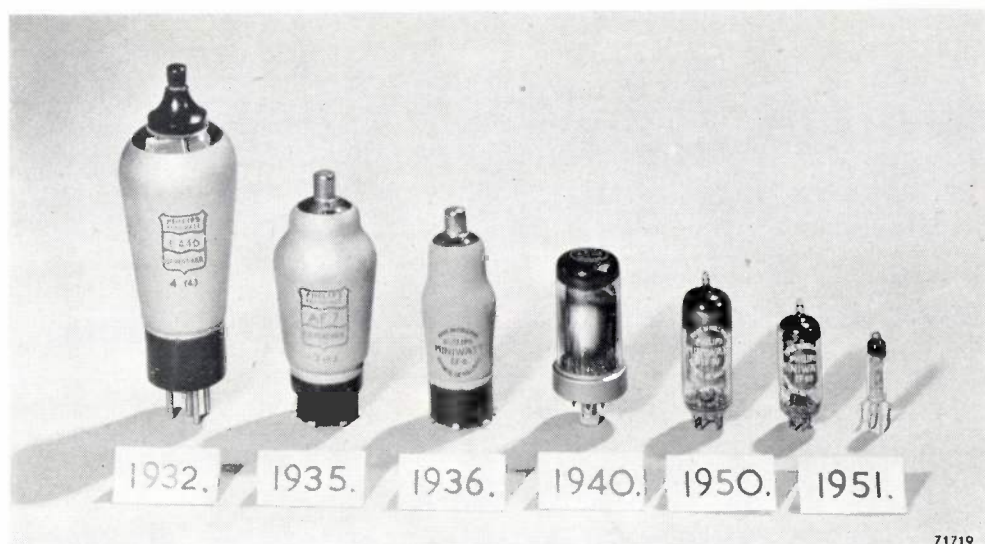


Fig. 3. In the course of time the dimensions of radio receiving tubes have been reduced again and again. All the tubes shown are pentodes.



Photograph Walter Nürnberg

Fig. 4. Apparatus employed for the measurement of secondary emission of solids. For a description see J. L. H. Jonker, The angular distribution of the secondary electrons of nickel, Philips Res. Rep. 6, 372-387, 1951 (No. 5).

of what occurs inside the tube and thus of the means of simplifying the computations⁶).

In the course of time endeavours to secure improved tube performance in the receiver have on many occasions shown the desirability of including more than one grid between cathode and anode, and in this way the well-known tetrodes, pentodes, hexodes, pentagrids, octodes, having respectively 2, 3, 4, 5 and 6 grids, originated. The record number of grids so far is found in the enneode (7 grids) employed as a phase detector for F.M. signals⁷). Small dimensions and a high degree of accuracy in the arrangement of the electrodes and their leading-in wires ensure that these broadcast

tubes will operate effectively at wavelengths down to a few metres⁸). The special form of construction necessary for tubes intended to work at still shorter wavelengths will be considered under a separate heading.

Whereas the configurations of the electrodes in tubes for the various functions in receivers have become more or less stabilized in recent years, the same cannot be said of tubes for television reception. The short wavelengths, the wide frequency band and the saw-tooth voltages required for the motion of the cathode ray over the screen impose other requirements on the tubes than those necessary for ordinary broadcast reception. These

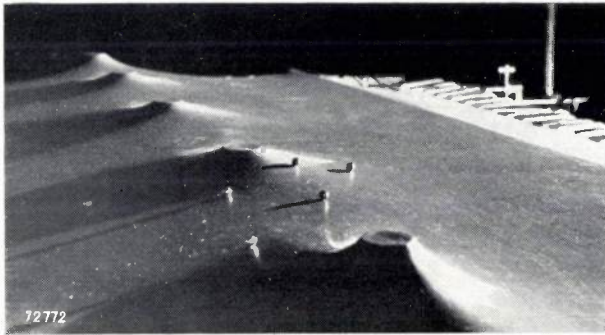


Fig. 5. This deformed rubber sheet reproduces the potential field in the cross section of a plane triode having 5 grid rods, for given values of the grid and anode potentials. Steel balls rolling over the sheet describe the same paths as electrons in the potential field.

requirements are again best met by tubes which are specially designed for the purpose⁹⁾. A similar situation will arise in the future in the case of colour-television receivers, and we have every justification for anticipating that new designs will be evolved which will render the receivers more sensitive and less complicated.

Efforts to reduce the number of tubes per receiver by introducing special types would at first sight appear to constitute a laborious, suicidal policy on the part of the manufacturer. This is by no means the case, for simpler and less expensive receivers can be sold in larger quantities. Apart from that, it is difficult to keep in check improvements and simplifications of techniques.

Transmitting tubes

In transmitting tubes, the main function of which is to supply considerable power with a high degree of efficiency, the maximum temperature

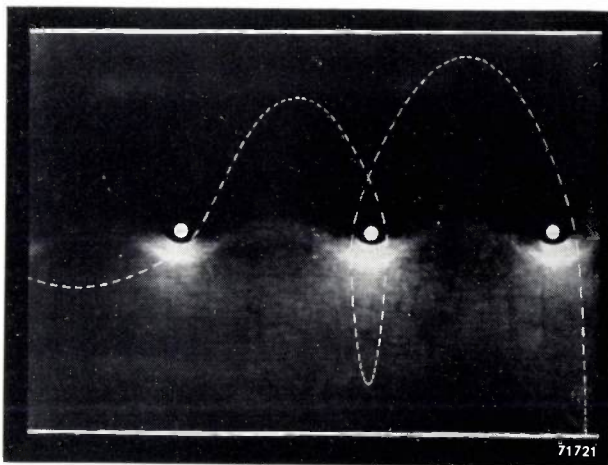


Fig. 6. Example of path of ball on the rubber sheet as photographed in intermittent light. The photograph shows that under certain conditions the electron can undergo such deflection between the grid wires that it is flung beyond the space between the electrodes.

which the tube materials are able to sustain determines the limit of utilization and, if such be possible, technology here plays an even more important part than in receiving tubes. Amongst classical types, i.e. the triode, tetrode and pentode, the volume of the tube for a given output power has in recent years been reduced by a factor of from 5 to 10 (fig. 7) by employing high-melting-point materials with good radiation properties for the electrodes, non-volatile getters and new types of glass having improved thermal and electrical characteristics, as

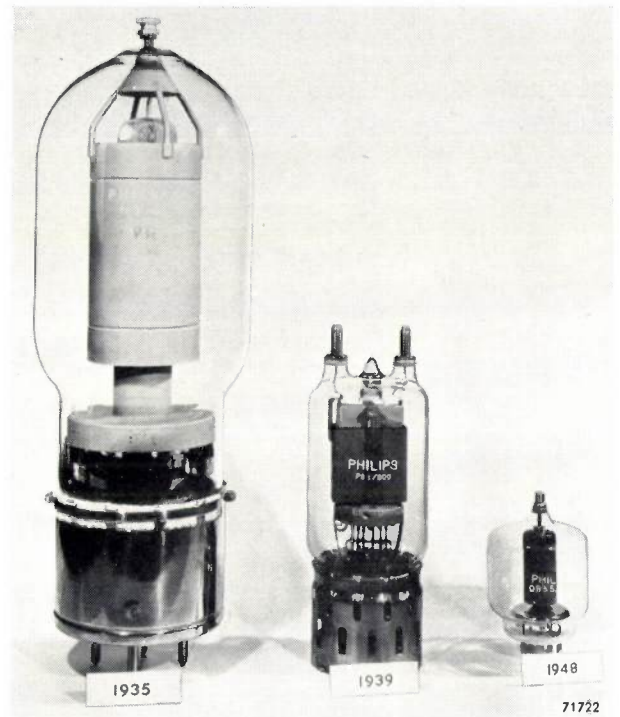


Fig. 7. Evolution of the transmitting tube with the years. All three types deliver about 1000 W; the latest of these is 14.5 cm in height.

well as by making a special study of the dimensioning of the electrodes and their leading-in wires¹⁰⁾. The reduction in inter-electrode capacitances brought about by thus curtailing the size of the tubes is particularly important at the very short wavelengths, where low capacitance is a *conditio sine qua non*. The maximum permissible loading of such small electrodes, however, sets a limit to the output power. When it is not possible with given electrode dimensions to dissipate enough of the heat by radiation, forced cooling with air or water is employed (fig. 8), for which purpose highly efficient cooling systems have been designed in which the most advantageous turbulence of the coolant is ensured¹¹⁾. In the future, especially among short-wave transmitting tubes, further



Fig. 8. Water-cooled transmitting tube for 100 kW output, with its anode water-cooling jacket.

improvements in consequence of continued technological research, as also further refinements in the electrode system, are to be anticipated.

Tubes for ultra-short waves

The urge towards the new and unknown has prompted many experimenters to investigate shorter and still shorter wavelengths. For the original radio links over great distances the long-wave range of 1000-20000 metres was employed, but techniques have in the course of time tended more and more towards shorter wavelengths for such purposes. As to this, we are much indebted to amateurs all over the world for the part they played in this pioneer work.

At present normal communication transmitters, amongst which those intended for broadcasting, operate at wavelengths ranging from 25 kilometres to about 1 metre, whilst for normal television wavelengths of several metres and for colour television decimetric waves are used. Centimetric waves are now used for radio relay stations and for radar; millimetric waves serve physical science in the study of the characteristics of molecules and atomic nuclei by means of the absorption spectroscopy of gases¹²).

The ordinary receiving tube is not effective at decimetric and shorter waves, when the frequency is increased to such an extent, the gain per stage drops below unity. The causes of this, viz. electrical losses, radiation and transit-time effects, are in part ascribable to the characteristics of the circuit at such wavelengths; in part, too, they are of electronic origin¹³). Here, other forms of tuning devices, as well as tubes of different design are needed to ensure success; for example, in the centimetric and millimetric wavelengths, the circuit consists of transmission lines, coaxial cables, wave-guides and cavity or rod-shaped resonators¹⁴), the radiation and I^2R losses of which can be made appreciably lower than those of the conventional types of coil employed for the longer wavelengths. It has been found possible to extend the working range of the tubes towards the shorter wavelengths, firstly by reducing the size and spacing of the electrodes so as to cut down capacitances and transit-time effects and, secondly, by making the leading-in wires for the electrodes function as part of the resonant circuits. The latter step has led to the use of disc-shaped or annular electrode connections in short-wave tubes; the metal discs, being thus part of the resonant circuit, are sealed laterally through the wall of the glass envelope, so that the part inside can serve as support for an electrode (*fig. 9*).

As to the reduction in the electrode spacing, it has in some cases been found possible to cut down the space between the control grid and cathode to



Fig. 9. "Disc-seal triode" type EC 56. Grid and anode are carried by metal discs sealed laterally into the glass wall; the outer part of the discs forms part of the resonant circuit.

as little as a few hundredths of a millimetre¹⁵). So that the controlling action shall not be lost in consequence of this, the mesh of the grid has to be unusually fine; to this end taut tungsten wires 10 microns in thickness are used, this being many times thinner than human hair (*fig. 10*). Transit times in an oscillator can be reduced still further by the use of higher voltages, which, in order not to overload the electrodes, are applied in pulses; the electrodes thus have time to cool between two pulses.

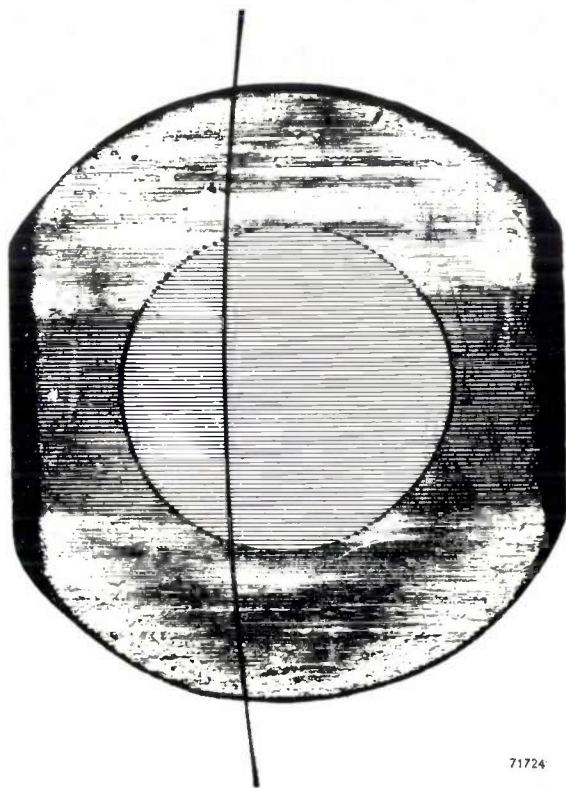


Fig. 10. Micro-photograph of the grid incorporated in the triode depicted in *fig. 9*. The thickness of the grid wires is roughly 10 μ . A human hair is shown across the wires for comparison. Magnification 10 \times .

Clearly there is a practical limit beyond which it is not possible to manufacture tubes with still smaller dimensions; it would seem improbable, therefore, that we shall be able to get any further than wavelengths of some centimetres by this means. Ultimately the limited loading capacity of the electrodes, as well as the impracticability to fix electrodes operating at elevated temperatures with extremely small spacings, will prove the deciding factors. Another limitation is found in the size of the cathode, since emission, and therefore also the output power, decrease with the area of the cathode. Increased emission through higher specific loading can be achieved only at the expense of the life. Under pulsed conditions, however, the specific

emission of the ordinary oxide-coated cathode can be considerably increased¹⁶).

In order to lengthen the life when the specific emission is increased, special cathodes have been developed, such as the L cathode in which the same alkaline earth oxides are used as for the conventional oxide-coated cathode, but applied behind a porous layer of sintered tungsten¹⁷).

If high power is required at very short wavelengths, it is necessary to resort to electrode systems whose dimensions are of the same order as the wavelengths at which these systems should operate; this, now, is the case in tubes whose action is based on a utilization of transit-time effects, as in magnetrons which during the second world war played such an important part in radar for locating submarines and aircraft, as also in klystrons, travelling-wave tubes and similar systems¹⁸). In such tubes the transit time of the electrons is of the same order as the periodic time of the alternating current; thus the electrons are able to impart energy to the high-frequency field which is thereby increased so that oscillations are produced. Although such systems do not excel by reason of their efficiency, they have the advantage that they are large enough with respect to the short wavelength not only to dissipate the heat generated, but to be manufactured without serious mechanical difficulties.

For generating waves of 3 and 10 cm, klystrons and magnetrons have been made that supply a peak output of respectively 10 and 5 megawatts, with efficiencies of 30% and 60%. On the other hand, a 3 cm magnetron has been manufactured for a peak output of 100 W which weighs not more than 1 kg including the permanent magnet (*fig. 11*).

The shortest wavelength for which magnetrons can be manufactured is in the region of a few millimetres, the efficiency being rather low. It has also proved possible to obtain a small amount of power, as higher harmonic of the fundamental wave, at a wavelength of roughly 1 millimetre¹⁹).

Klystrons have been made for a continuous oscillation at a wavelength of a few millimetres, but, as the circuit losses are inversely proportional to the square root of the wavelength, the efficiency at such wavelengths is very low.

These tubes are accordingly not suitable for amplification purposes. The travelling-wave tube is much less susceptible to poor circuital properties in this range of frequencies, as it operates without a resonant circuit.

In these tubes a beam of electrons, usually concentrated by means of a magnetic field, is shot in axial direction through an elongated wire helix,

the signal being applied to the cathode side of the helix. When the speed of propagation of the beam and that of the field of the travelling wave in the axial direction are made almost identical, the signal is amplified along the length of the helix, at the expense of the beam. With greater losses in the helix it is only necessary to employ a longer helix for the same amount of gain.

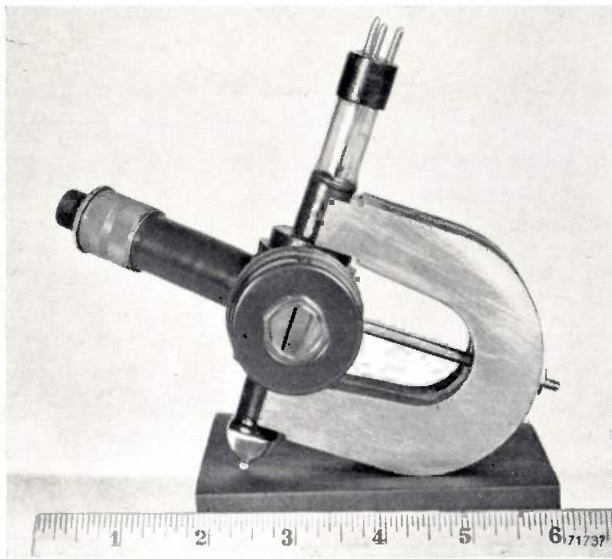


Fig. 11. Magnetron for a wavelength of 3 cm designed by Philips Laboratories, Irvington, N.Y., U.S.A. This magnetron is not intended for radar, but for a radio beacon. Peak output 100 W at an anode voltage of 800 V, heater power 2 W. Total weight including the permanent magnet: approx. 1 kg. The size may be seen from the inch rule placed in front of the magnetron. (Photo by courtesy G. A. Espersen and B. Arfin, *Tele-Tech.* 10, No. 6, p. 50 and No. 7, p. 30, 1951.)

On this principle tubes have been made that will provide ample gain at a wavelength of only a few centimetres, and the success achieved with these tubes on such wavelengths has stimulated experiments with numerous similar systems, in which the helix is replaced by rods, wires, a wave-guide, or a second electron beam.

There is still an enormous amount of scope for microwave experimentation, if only by reason of the fact that among the many different methods of making a high-frequency alternating field and a constant electric and magnetic field cooperate spacially with a beam of electrons, only few have been investigated in tubes. More systematic research than has hitherto taken place might yield remarkable possibilities.

In many laboratories all over the world efforts are being made along widely divergent lines to penetrate further into the interesting problem of generating and amplifying microwaves.

Cathode-ray tubes

In cathode-ray tube use is made of a narrow beam of high-speed electrons to produce a luminous image. This is achieved by deflecting the beam, or cathode ray, electrically, so that it travels over a surface coated with substances that emit light in consequence of the electron bombardment. Originally these tubes were employed in the cathode-ray oscilloscope. The coming of television has imposed new conditions on the quality, and mass production, too, has brought other considerations to bear. A distinction is made between two systems in television, viz. direct and indirect vision of the images. Until now the former has been the more widely used of the two systems, but this has led to the making of tubes with larger and larger screens, seeing that small images tend to tire the viewer and to limit the number of viewers per receiver. In order to reduce the weight of such large tubes, the heavy glass cone has been replaced by a lighter one of metal²⁰) and the inconveniently large dimensions have been cut down by introducing a rectangular screen to take the place of the round one (see also frontispiece), and shorter electrode systems for generating and deflecting the electron beam. It has also been found possible to shorten the cone by employing a narrower electron beam which is capable of deflection through larger angles (90°) (fig. 12).

It may well be asked whether it is desirable to go any further in this direction to produce still larger pictures. The presence in the living room of a vacuum tube 75 cm in diameter, as recently produced in the United States, would seem to be anything but desirable, even though the danger of implosion due to the atmospheric pressure of roughly 4 tons on the glass screen need not be considered high in a properly designed tube. In order to withstand such forces, the screen has to be very thick, and also more or less convex, which means a certain amount of distortion for the viewer. Lastly it may be said that a tube of such dimensions presents quite a problem for the set manufacturer in the design of a cabinet that will be suitable for installing in the living room.

If large or very large pictures are demanded, the indirect system is much to be preferred; a cathode-ray tube of quite small dimensions but very high luminous intensity then provides an image that can be projected on to a flat screen by means of lenses or a concave mirror²¹). The high luminous intensity of the screen required in this case imposes very stringent conditions on the luminescent substances from the point of view of loading.

Camera tubes for television

Developments in camera tubes for television have led to magnificent and most interesting results. We have been given the iconoscope, the image iconoscope (*fig. 13*), the orthicon and the image orthicon, associated with such names as Zworykin, Rose and Iams²²). These tubes, which convert the optical image into an electrical signal and which rank amongst the most complex of electronic inventions, were the means of appreciably hastening forward the realization of practical television. The technological problems to be solved were great. For

per second by a beam of electrons, the necessary signal being obtained by "measuring" the charges of the photo-cells along those lines.

Means have been found of ensuring a very high sensitivity to light in the image orthicon, namely 10 times higher than that of the fastest photographic material; hence a serviceable image is obtained by the light of a candle, or by moonlight. This has been achieved, firstly, by using for the scanning a beam of electrons having a comparatively low velocity, thus eliminating the interfering secondary emission from the mosaic inherent in the

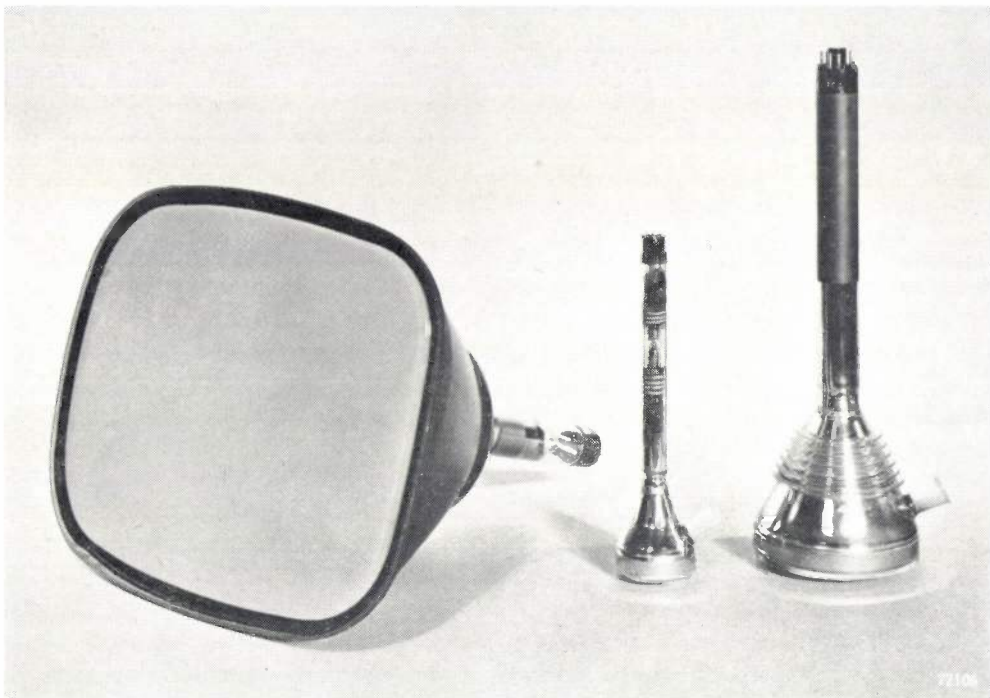


Fig. 12. Cathode-ray tubes for television. At the left a rectangular direct-vision tube with metal cone giving a picture 28 cm \times 37 cm. Centre: tube for projection television in the home (max. 1.0 m \times 1.2 m approx.). At the right: a tube for projection in halls (3 m \times 4 m).

instance in the iconoscope the optical image is translated into a pattern of electrical charges by means of a screen or mosaic about 100 sq. cm containing several milliards of minute photo-cells, each insulated from its neighbours, in a density of roughly 360 000 per sq. mm. For comparative purposes it may be pointed out that the concentration of the rods and cones in the retina of the human eye is at most 20 000 per sq. mm, which means that the artificial retina of the iconoscope improves on nature to the extent of almost 20 times (*fig. 14*). In order that the image may be transmitted, it is scanned in lines at a speed of thousands of metres

iconoscope and, secondly, by making effective use of secondary emission by including in this already complicated tube a multi-stage electron multiplier.

Thus, in order to ensure distortionless conversion of the optical image into electrical signals in these tubes, a number of electronic mechanisms, each of which is already sufficiently complicated, must be made to work together in the correct manner, which means that the manufacture of this kind of tube is certainly no sinecure. As with other types of tube, endeavours are being made to render the camera tube a more compact instrument: amongst other things, this will admit of more convenient

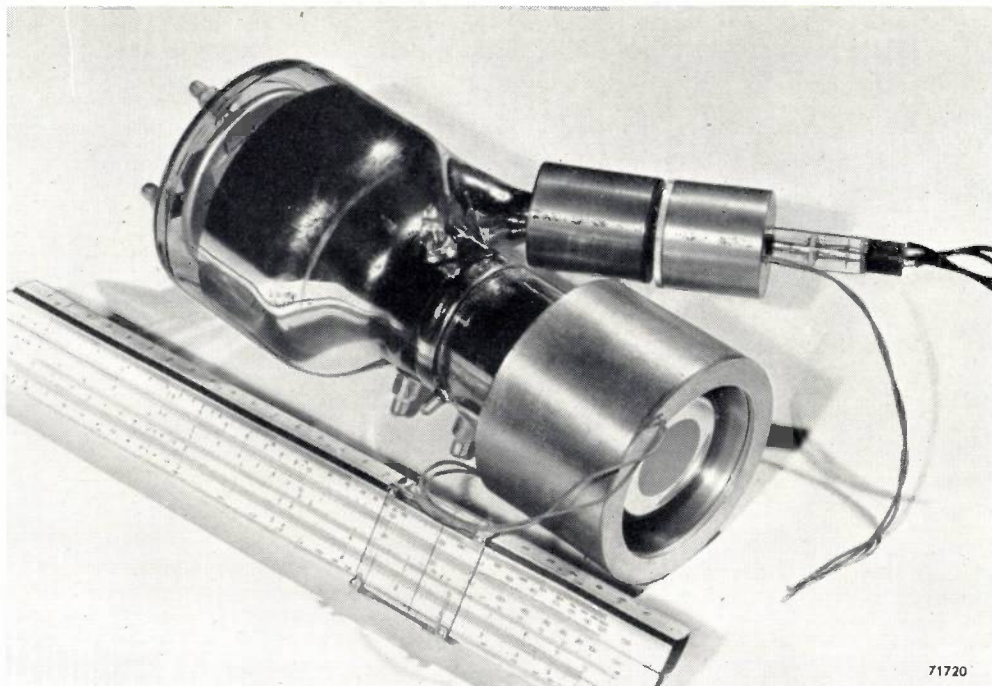


Fig. 13. Image iconoscope. At the near end of the tube will be seen the photo-cathode on which the scene to be televised is reproduced by means of a photographic lens. The large magnet coil shown at this end of the tube produces the field required to focus the image from the photo-cathode on to the mica target, of which a part is just visible at the opposite, wide, end of the tube. At the far side the neck with electron gun and deflection coils for producing the electron beam which periodically scans the target.

proportioning of the optical system, increased depth of focus and less expensive lenses. It is not unlikely that in the future higher sensitivity in this kind of tube will be sought in a wider use of photo-conductivity in place of photo-emission, seeing that the quantum efficiency of the former is in excess of unity. The phenomenon of induced conductivity in insulators also has possibilities to offer in the design of camera tubes²³).

In conclusion

Only a few examples from the very extensive field of electronic tubes have been discussed here: other important applications, such as those of counter and "memory" tubes, gas-filled tubes, switching tubes, tubes for physical research or for industrial uses and many others have been omitted from our review. In most of these

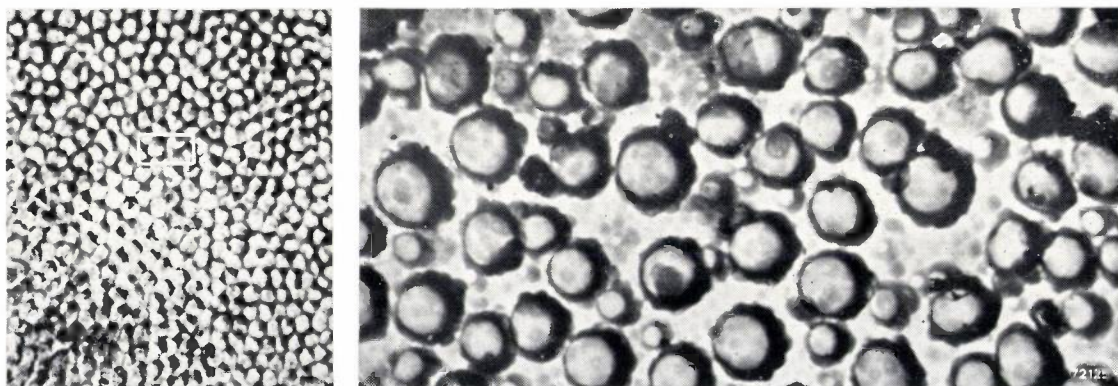


Fig. 14. *a*) Photograph of part of the retina of the human eye magnified 750 times. (From S. L. Polyak, *The Retina*, Univ. Chicago Press, Chicago 1941).
b) Part of the mosaic of an iconoscope, photographed with the electron microscope. Magnification 11 000 \times . The part shown has the same size as that area of the retina enclosed within the small marked rectangle in photograph (*a*).

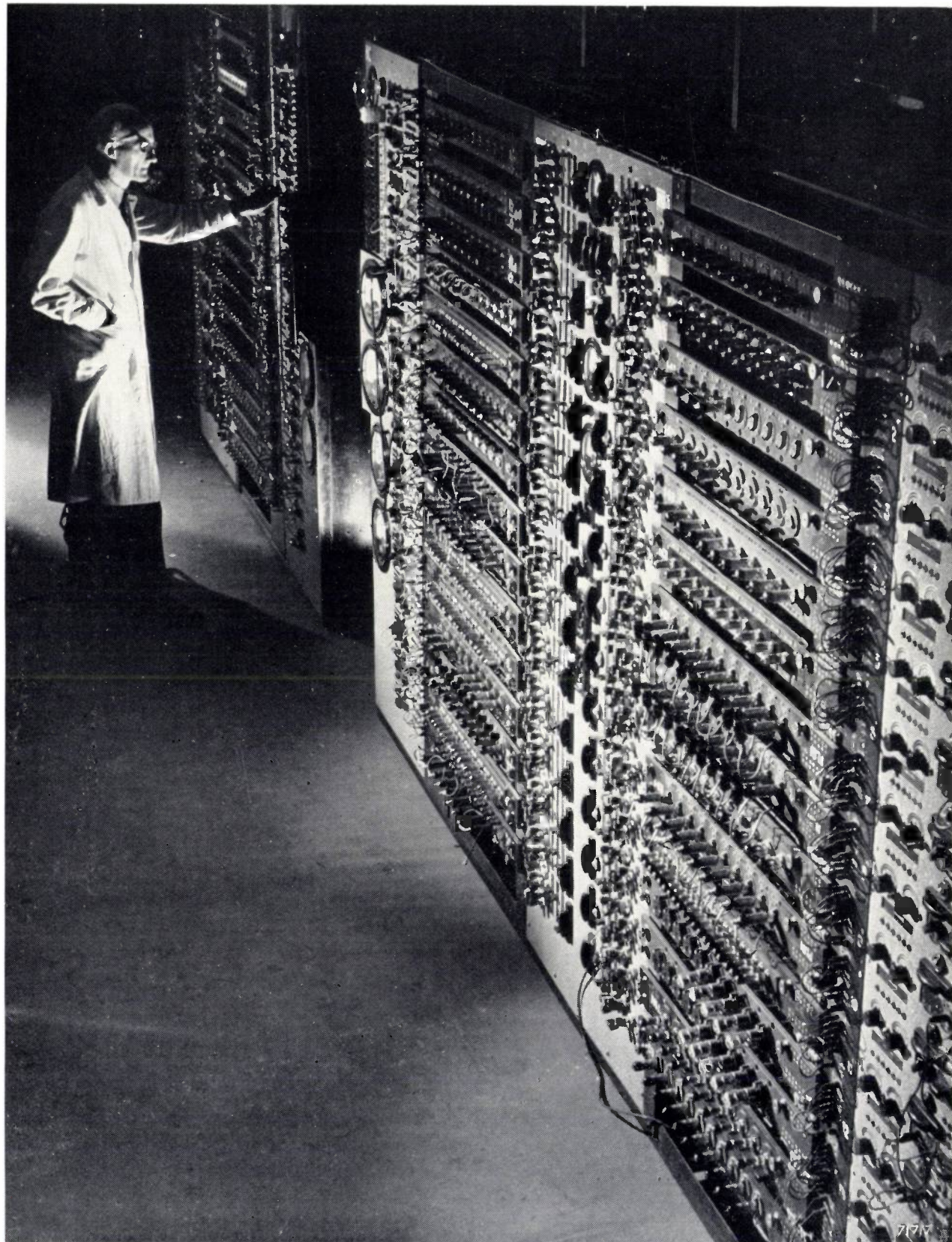
categories exploratory work is still in full progress and new departures are of almost daily occurrence. The limits of the possibilities offered by electronic tubes are still invisible; even after half a century of research and development, fresh possibilities are continually being discovered.

BIBLIOGRAPHY

The references added to Prof. Jonker's address make no pretence at being a comprehensive review or even a representative cross section of existing literature on the subject dealt with. In this selection the work of the Physical Laboratory at Eindhoven has naturally been placed somewhat in the foreground.

- 1) R. C. Coile, Proc. I. R. E. **38**, 1380-1384, 1950.
- 2) U. S. Patent No. 879-532, applied for 29 Jan. 1907, granted 18 Febr. 1908.
- 3) Franklin Inst. State Penn. for the Prom. Mech. Arts Comm. Sci. Arts, Rep. No. 3087, 8 Jan. 1941, p. 3-4.
- 4) M. Chodorow, E. L. Ginzton, I. Neilsen and S. Sonkin, I. R. E. National Convention, New York, March 1950.
- 5) J. L. H. Jonker, Wirel. Eng. **16**, 274-286, 344-349, 1939; A. J. Heins van der Ven, Wirel. Eng. **16**, 383-390, 444-452, 1939.
- 6) C. L. Fortescue and S. W. Farnsworth. Proc. Amer. Inst. El. Eng. **32**, 757-772, 1913 (principle of electrolytic tank).
G. Hepp, Philips techn. Rev. **4**, 223-230, 1939 (description of electrolytic tank).
P. H. J. A. Kleijnen, Philips techn. Rev. **2**, 338-345, 1937 (method of rubber sheet).
For application and results of the methods see, i.a., J. L. H. Jonker, Philips techn. Rev. **5**, 131-140, 1940; J. L. H. Jonker and B. D. H. Tellegen, Philips Res. Rep. **1**, 13-32, 1945;
J. L. H. Jonker, Philips Res. Rep. **1**, 331-338, 1946; **4**, 357-365, 1949; **6**, 1-13, 1951.
- 7) J. L. H. Jonker and A. J. W. M. van Overbeek, Philips techn. Rev. **11**, 1-32; 1949.
- 8) F. Prakke, J. L. H. Jonker and M. J. O. Strutt, Wirel. Eng. **16**, 224-230, 1939; G. Alma and F. Prakke, Philips techn. Rev. **8**, 289-295, 1946.
- 9) R. M. Cohen, R. C. A. Rev. **12**, 3-25, 1951.
- 10) For a survey see, e.g., J. P. Heyboer, Transmitting valves, revised by P. Zijlstra, Philips' Technical Library, 1951.
- 11) H. de Brey and H. Rinia, Philips techn. Rev. **9**, 172-178, 1947 (air cooling); M. J. Snijders, Philips techn. Rev. **10**, 239-246, 1949 (water cooling).
- 12) Summarizing articles on this subject: W. Gordy, Rev. mod. Phys. **20**, 668-717, 1948; D. K. Coles, Adv. in Electronics **2**, 299-362, 1950.
- 13) M. J. O. Strutt and A. van der Ziel, Proc. I. R. E. **26**, 1011-1032, 1938.
- 14) See, e.g., C. G. A. von Lindern and G. de Vries, Philips techn. Rev. **6**, 217-224, 1941. (box-shaped resonators); **6**, 240-249, 1941 (transmission lines); **8**, 149-160, 1946 (flat cavity resonators). G. de Vries, Philips techn. Rev. **9**, 73-84, 1947 (cavity resonators). W. Opechowski, Philips techn. Rev. **10**, 13-25 and 46-54, 1948 (wave-guides). S. Millman, Proc. I. R. E. **39**, 1035-1043, 1951 (rod-type resonators).
- 15) J. A. Morton and R. M. Ryder, Bell Syst. techn. J. **29**, 496-530, 1950.
- 16) E. A. Coomes, J. appl. Phys. **17**, 647-654, 1946; M. A. Pomerantz, Proc. I. R. E. **34**, 903-910, 1946.
- 17) H. J. Lemmens, M. J. Jansen and R. Loosjes, Philips techn. Rev. **11**, 341-350, 1950.
- 18) On magnetrons see: J. B. Fisk, H. D. Hagstrum and P. L. Hartman, Bell Syst. techn. J. **25**, 167-348, 1946; brief survey by J. Verweel, Philips techn. Rev. **14**, 44-58, 1952 (No. 2).
On klystrons see, e.g., D. R. Hamilton, J. K. Knipp and J. B. H. Kuper, Klystrons and microwave triodes, McGraw-Hill, New York 1948; also B. B. van Iperen, Philips techn. Rev. **13**, 209-222, 1952 (No. 8).
On travelling-wave tubes see: R. Kompfner, Wirel. Eng. **24**, 255-266, 1947; J. R. Pierce and L. M. Field, Proc. I. R. E. **35**, 108-111, 1947; also B. B. van Iperen, Philips techn. Rev. **11**, 221-231, 1950.
For other systems see, e.g., L. M. Field, Electronics **23**, Jan. 1950, 100-104.
- 19) J. A. Klein, J. H. N. Loubser, A. H. Nethercot Jr. and C. H. Townes, Rev. sci. Instr. **23**, 78-82, 1952 (No. 2).
- 20) H. P. Steier, J. Kelar, C. Y. Lattimer and R. D. Faulkner, R. C. A. Rev. **10**, 43-58, 1949.
- 21) P. M. van Alphen, H. Rinia, J. de Gier, G. J. Siczen, F. Kerkhof and J. Haantjes, Philips techn. Rev. **10**, 69-78, 97-104, 125-134, 307-317 and 364-370, 1948/49.
- 22) V. K. Zworykin, Proc. I. R. E. **22**, 16-32, 1934 (iconoscope); H. Iams, G. A. Morton and V. K. Zworykin, Proc. I. R. E. **27**, 541-547, 1939 (image iconoscope); A. Rose and H. Iams, R. C. A. Rev. **4**, 186-199, 1939 (orthicon); A. Rose, P. K. Weimer and H. B. Law, Proc. I. R. E. **34**, 424-432, 1946 (image orthicon).
See also P. Schagen, H. Bruining and J. C. Francken, Philips techn. Rev. **13**, 119-133, 1951 (No. 5).
- 23) P. K. Weimer, S. V. Fergue and R. R. Goodrich, R. C. A. Rev. **12**, 306-313, 1951 (photo-conductivity); L. Pensak, Phys. Rev. **75**, 472-478, 1949 (induced conductivity).

LIFE TESTS IN THE ELECTRON TUBE FACTORY



Photograph Walter Nürnberg

THE IMPULSE-GOVERNED OSCILLATOR, A SYSTEM FOR FREQUENCY STABILIZATION

by E. H. HUGENHOLTZ *).

621.396.615.12:621.316.726

The introduction of the quartz crystal in radio technique 25 years ago solved the problem of constructing an oscillator with constant frequency. A drawback of the crystal, however, is that it is suitable for only one frequency so that transmitters which are to operate at different frequencies, require several crystals with the present-day circuit designs, and in a transmitter operating with a continuously variable frequency the crystal cannot be used at all.

This article describes a method by which, with only one crystal, signals can be generated with a series of very stable frequencies, and, if required, a signal can be obtained with a continuously variable frequency which is nevertheless very stable.

Introduction

Two important problems have arisen in relation to the frequency of the oscillators now used in radio transmitters and receivers because of increased traffic and speed. These are:

- a) it must be possible to adjust the frequency very accurately;
- b) there must be absolute stability of the frequency under changing working conditions, for instance, the ambient temperature or the supply voltages of the valves.

In addition, it is sometimes necessary to change the frequency at short notice.

In transmitters operating at one, constant, frequency and in which frequency changes are comparatively rare, as is the case in broadcasting transmitters, a quartz crystal¹⁾, usually mounted in a thermostatically controlled container, is generally used. Quartz crystals can also be used in transmitters which use several frequencies for broadcasting, which can be selected as the necessity arises. In that case there must be a crystal for each frequency, and the thermostatically controlled container in which the crystals are mounted must be large. A drawback of this system is that it takes a fairly long time before the required temperature is reached, which may lead to difficulties in mobile installations. The cost of a great number of quartz crystals is another deterrent.

If the frequency of an oscillator must be continuously variable in a certain range, an oscillatory circuit with a variable capacitor is generally used. Various measures are then taken to ensure that the oscillator frequency depends as little as possible upon the operating conditions. Compensation is applied for frequency variations caused by tempera-

ture fluctuations, and considerable accuracy of adjustment can be achieved by dividing the required frequency range into a number of smaller ranges which can be selected by changing coils. This often makes the installation fairly complicated, however, and the usual requirements of direct-reading dials cannot be easily met. Moreover, the frequency stability that can be obtained is always less than that which can be achieved with a quartz crystal.

If the frequency range must be continuously variable only in a limited range, a high degree of stability can be achieved with the type of circuit

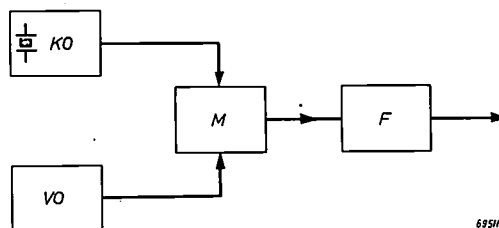


Fig. 1. Fundamental diagram of a system for obtaining an alternating voltage whose frequency can be continuously varied within a limited range and possesses considerable stability.

KO = crystal-controlled oscillator; VO = oscillator of which the (continuously variable) frequency is much lower than that of KO; M = mixer; F = filter.

of which the principle is given in *fig. 1*. If, for example, a voltage is required having a frequency that can be varied between 1 and 1.1 Mc/s, it may be obtained by supplying two voltages to a mixer valve *M*. One of these is supplied by a crystal-controlled oscillator *KO*, having a frequency of, for example, 0.8 Mc/s, and the other by an oscillator *VO*, of which the frequency is continuously variable within a range of 0.2 to 0.3 Mc/s. The mixer supplies, amongst others, a voltage at a frequency which is equal to the sum of the two oscillator frequencies

*) N.V. Philips' Telecommunicatie Industrie, Hilversum.

¹⁾ See, for instance, J. C. B. Missel, Piezo-electric materials, Philips techn. Rev. 11, 145-150, 1949.

and which can be varied between 1 and 1.1 Mc/s. As the frequency changer also supplies voltages with frequencies which are not required, the voltage required will have to be supplied via a filter F .

As the oscillator VO operates at a comparatively low frequency, a greater frequency stability can be obtained than by using an oscillator having a frequency that is continuously variable between 1 and 1.1 Mc/s. It will be clear, however, that this advantage is no longer present when a voltage is required with a continuously variable frequency over an extensive range, for example, from 1 to 10 Mc/s. This range could be achieved, of course, if a large number of crystal oscillators were available with frequencies of, for example, 0.8 Mc/s, 0.9 Mc/s, 1.0 Mc/s, 1.1 Mc/s etc., and, in addition, an oscillator with a continuously variable frequency in a limited range, for instance, from 0.2 to 0.3 Mc/s. This solution is obviously not very attractive.

The system described becomes impracticable at higher frequencies when the ratio between the crystal frequency and the frequency of the continuously variable oscillator will generally be large. This means that the relative distance between the required frequency and the unwanted crystal frequency is small, so that high requirements are imposed on the filter F which is intended, in the first place, to suppress the crystal frequency (and also the image frequency).

Another possibility is to use one crystal oscillator with a frequency of, for example, 0.1 Mc/s and to obtain from it, by means of frequency multiplication, voltages of which the frequency is a whole multiple of the oscillator frequency. Frequency multiplication may be obtained, for example, by means of a so-called impulse generator, i.e., a generator which does not produce a sinusoidal alternating voltage, but a series of periodic voltage impulses. Such an impulse generator can be synchronized with the crystal oscillator, so that the frequency of the impulses is equal to the frequency of the oscillator voltage. It is possible to obtain from the periodic voltage impulses, by means of filters, alternating voltages of which the frequency is a whole multiple of the impulse frequency. This system, however, requires a complicated set of filters, the very high selectivity requirements of which can hardly be realized, particularly if high ordinals of harmonics are required.

The following will describe another solution of this problem, namely the Impulse-Governed Oscillator (IGO). This method enables the frequency of an oscillator to be adjusted, quite simply, to a large number of values with the same degree of

accuracy as would be achieved if a quartz crystal were used for each of these frequencies.

The principle of the IGO is frequency stabilization by means of an automatic control system (often referred to as the "automatic monitor system"), which will be explained first in more detail.

Stabilizing an oscillator frequency by means of an automatic control system ("automatic monitor system")

Fig. 2 shows the principle of this method of frequency control. The stabilized oscillator SO supplies a signal of frequency f_1 which is applied to the mixer M . A signal of another frequency, f_2 , supplied by an oscillator KO , and of which the frequency is stabilized by means of a quartz crystal, is simultaneously

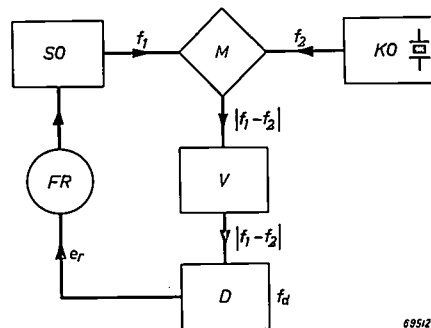


Fig. 2. Fundamental diagram of the "automatic monitor system" for frequency stabilization. SO = controlled oscillator, KO = crystal-controlled oscillator, M = mixer, V = amplifier, D = discriminator (frequency detector), FR = frequency control device.

supplied to the mixer. The mixer M supplies, amongst others, an alternating voltage of frequency $|f_1 - f_2|$ and this voltage is amplified in the amplifier V , after which it is fed to the discriminator (frequency detector) D . The circuits of the discriminator are tuned to a certain frequency f_d . The discriminator supplies a direct voltage e_r which is either positive or negative according to whether $|f_1 - f_2|$ is greater or smaller than f_d . If $|f_1 - f_2|$ is equal to f_d , then $e_r = 0$. The direct voltage e_r , which is called the control voltage, is fed to a control device FR which enables the oscillator frequency to be varied. This control device may consist, for example, of a capacitor whose shaft is driven by a small electric motor, or it may consist of a direct control device, for example a reactance valve²⁾. The arrangement of FR may be such that, when e_r is positive, the frequency of SO is reduced and vice versa. Thus, FR always influences the frequency of SO to such an

²⁾ See, for example, Th. J. Weijers, Frequency modulation, Philips techn. Rev. 8, 42-50, 1946.

extent, that $|f_1 - f_2|$ is approximately equal to f_d . The frequency f_1 of the oscillator SO is, therefore, approximately stabilized to the value $f_2 - f_d$ or $f_2 + f_d$, depending on f_2 being larger or smaller than f_1 . It is easy to see, however, that this frequency control will never achieve that $|f_1 - f_2|$ is exactly the same as f_d , for in that case the control voltage e_r would be zero.

If the discriminator D supplies a control voltage e_r of a volt per kc/s deviation between $|f_1 - f_2|$ and f_d , and if the detuning to which the oscillator SO is subjected by the control device FR is equal to b kc/s per volt control voltage, each deviation of f_1 from the required value $f_2 \pm f_d$ would be reduced by a factor $a \cdot b$ ³⁾.

The discriminator D may be constructed in the same way as the frequency detectors used in F.M.

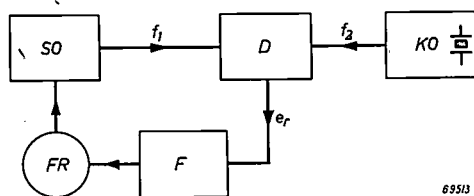


Fig. 3. Fundamental diagram of a system for frequency control. If D is a comparative discriminator, a control with proportional action is obtained; if D is a phase discriminator, frequency control with integral action is obtained. The significance of the other letters is given in figs 1 and 2.

receivers²⁾. The frequency stability of the oscillator SO is determined, on the one hand, by that of the crystal oscillator KO and, on the other hand, by the stability of the discriminator D . Generally speaking, f_d is chosen much lower than f_1 and f_2 , so that the absolute value of the residual deviation between $|f_1 - f_2|$ and f_d is small in relation to f_1 . The error in the frequency of SO , caused by the fact that there will always remain a control voltage e_r , however, usually considerably exceeds the sum of the errors of KO and D .

The same objection is found in another control system of which the principle is shown in fig. 3. In this system the voltages supplied by SO and KO are applied to a so-called comparative discriminator D , which is a circuit supplying a control voltage e_r of which the amplitude and sign depend upon the magnitude and the sign of the frequency difference between SO and KO ⁴⁾. Here too, the control voltage is supplied via a low-pass filter F

to the frequency control device FR , which influences the frequency of SO . It will be clear that in this case, too, equality of the SO and KO frequencies cannot be obtained.

It has been explained in the article quoted in footnote³⁾ that this disadvantage is inherent in every automatic control system with proportional action. With the aid of a control system with integral action, however, the deviation of f_1 with respect to f_2 can be reduced to nothing. To bring about such a control, D in fig. 3 must be replaced by a so-called phase discriminator⁵⁾. This refers to a circuit which supplies a control voltage e_r of which the amplitude and sign depend upon the magnitude and sign of the phase difference between the voltages supplied by SO and KO (thus the frequencies f_1 and f_2 are assumed to be equal in this case). As the frequency is the derivative of the phase²⁾, this is a control with integral action for the frequency.

If the frequencies f_1 and f_2 are not equal, D will supply an alternating voltage of which the frequency is equal to $|f_1 - f_2|$. If this frequency is lower than the cut-off frequency of the low-pass-filter F , this alternating voltage is supplied to the frequency control device FR , which affects the frequency of SO . Provided the inertia of FR is not too large, the frequency of SO will then vary periodically. When the value f_2 is passed, D supplies a direct voltage, which influences the frequency of SO , via F and FR , to such an extent, that this synchronization condition between SO and KO is maintained. In that case, there is no longer any frequency difference between SO and KO , but merely a phase shift.

If the frequency of SO still differs so much from that of KO that the difference is larger than the cut-off frequency of F , the frequency stabilization of SO will be non-effective. The "collecting zone", i.e., the frequency range in which the frequency of SO should be before stabilization begins to act, is, therefore, determined by the cut-off frequency of the filter F . The so-called "holding zone" is generally understood to be the frequency range within which SO can be detuned (once synchronization has taken place) before synchronization is lost. This frequency range is determined by the range in which FR can correct the frequency of SO and by the voltage supplied by the phase discriminator D , at a certain phase shift between the voltages supplied by SO and KO . Generally speaking, the "holding zone" is larger than the "collecting zone"

³⁾ H. J. Roosdorp, On the regulation of industrial processes, Philips techn. Rev. 12, 221-217, 1951. This concerns a so-called frequency control with proportional action.

⁴⁾ In order to make a distinction from this, a discriminator, as incorporated in the diagram of fig. 2 and to which only one alternating voltage is applied, is called an absolute discriminator.

⁵⁾ Also discussed in Philips techn. Rev. 13, 316 (fig. 4), 1952 (No. 11).

Fig. 4 is the fundamental diagram of a system for frequency control with integral action, whereby the frequency f_1 of the controlled oscillator *SO* can be varied over a limited range. This frequency is not stabilized to the frequency f_2 of *KO*, but to a

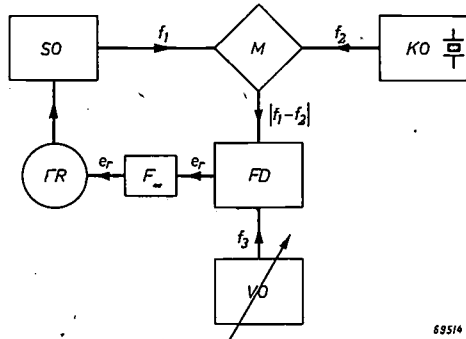


Fig. 4. Fundamental diagram of a frequency control system with integral action, in which there is a difference between the frequencies of the stabilized oscillator *SO* and the crystal-controlled oscillator *KO*, determined by the frequency of the oscillator with variable frequency *VO*. *FD* = phase discriminator. The significance of the other letters is as given in figs 1 and 2.

value which differs by a certain amount from f_2 . To achieve this, the voltages supplied by *SO* and *KO* are again supplied to the mixer *M*, which will supply a voltage having a frequency $|f_1 - f_2|$. This voltage is applied to a phase discriminator *FD*, together with the voltage from an oscillator *VO* of which

f_1 will be corrected by the same amount. If f_3 is chosen much lower than f_1 (in other words, if f_1 differs but little from f_2), a much larger stability can be obtained, than if *SO* itself were to be constructed as an oscillator with variable frequency.

By making a change in the scheme shown in the diagram in fig. 3, arrangements can be made whereby the oscillator *SO* is synchronized to a harmonic of the voltage of the crystal oscillator *KO*, so that the frequency of *SO* is exactly the same as a whole multiple of the frequency of *KO*.

An impulse generator is used for this purpose, so that such a circuit is known as an Impulse-Governed Oscillator (IGO) ⁶⁾.

The Impulse-Governed Oscillator (IGO)

The principle of the IGO is explained with the aid of the diagram of fig. 5. The valve B_1 with the oscillatory circuit $L_1 - C_1$, forms an oscillator (Hartley circuit). The frequency of the voltage supplied by this oscillator is influenced by the reactance valve B_2 as well as by L_1 and C_1 . An alternating voltage with the oscillator frequency which lags in phase about 90° with respect to the alternating anode voltage, is supplied to the grid of this valve via the phase-shifting network $R_2 - C_2$. This causes the anode current of B_2 to be about 90° lagging in phase with respect to the anode voltage, and the impedance of this valve, between anode and cathode, will

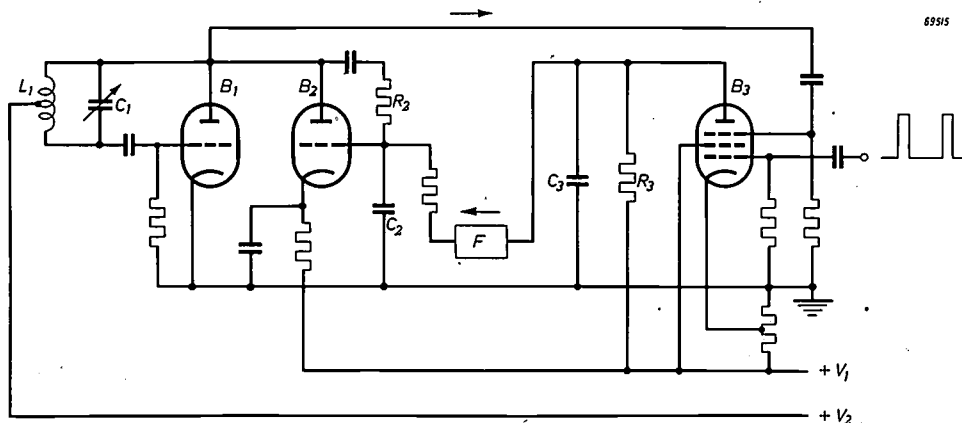


Fig. 5. Circuit diagram of an impulse-governed oscillator. B_1 = oscillator valve, B_2 = reactance valve, B_3 = impulse mixer valve.

the frequency f_3 can be varied within a certain frequency range. The voltage e_r supplied by *FD* is fed via the low-pass filter *F* to the frequency control device *FR*, thus influencing the frequency of the oscillator *SO*. The latter frequency will now be stabilized so that $|f_1 - f_2| = f_3$, thus f_1 will be equal to $f_2 + f_3$ or $f_2 - f_3$. If the frequency of *VO* is varied,

have a reactive character. The extent of the equivalent reactance depends, amongst other things, upon the mutual conductance of the valve and hence may be influenced by the grid voltage of B_2 .

⁶⁾ See also H. B. R. Boosman and E. H. Hugenholtz, Frequency control in transmitters, Communication News 9, 21-32, 1947.

The alternating voltage supplied by the oscillator is fed to a so-called impulse mixer valve, the pentode B_3 . In this valve, such a large positive voltage is applied to the cathode with respect to the first grid, that no anode current flows in the valve. The first grid is supplied with a series of periodic positive voltage impulses, obtained from an impulse generator which will be described later. The peak voltage of these impulses has such a high value that, during each impulse, anode current flows for a very short time in B_3 . The voltage supplied by the oscillator B_1 is applied to the third grid of B_3 .

The operation of a circuit in accordance with fig. 5 will be explained with the aid of fig. 6. This diagram shows the alternating voltage V_{g3} supplied by the oscillator and the voltage impulses V_{g1} applied to the first grid of B_3 . If the frequency of V_{g3} is a whole multiple of that of V_{g1} , the anode current impulses will always occur at the same phasing of V_{g3} .

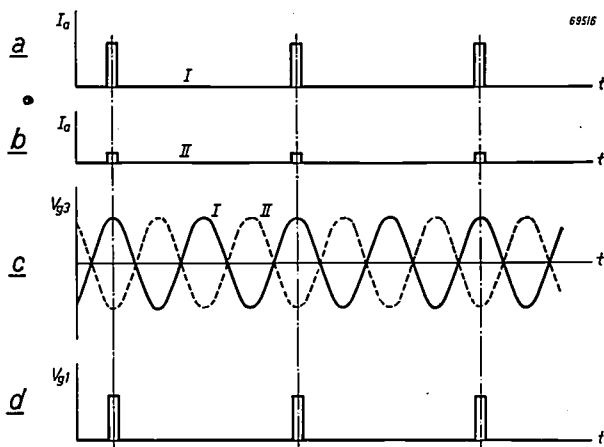


Fig. 6. Graphs explaining the operation of the IGO. The oscillator voltage V_{g3} at the third grid (c) and the impulse voltage V_{g1} at the first grid of B_3 are presented as functions of time (see fig. 5). The anode current impulses, for the case that the relative phasing of V_{g1} and V_{g3} , is as indicated by I and II (a and b resp.), are also indicated as functions of time.

If this is the position indicated in fig. 6c by I, the amplitude of the anode current impulses will be at its highest, in II it is at its lowest. These current impulses cause a periodical charging of the capacitor C_3 (fig. 5). The capacitance of this is so large that a practically ripple-free direct voltage occurs across C_3 , and the amplitude of this depends upon the relative phasing of V_{g1} and V_{g3} . This voltage is applied via the low-pass-filter F to the grid of the reactance valve B_2 . The average value of this control voltage is obtained when the impulses at the first grid occur at exactly that moment when the alternating voltage V_{g3} passes through zero. Thus, a very small deviation in this setting causes

a larger or smaller control voltage to occur. This direct voltage, in turn, influences the frequency of the oscillator, through the reactance valve B_2 . If the tuning of this oscillator is changed by an amount which is not too large, the phasing of V_{g3} will change with respect to that of V_{g1} , causing the control voltage to be altered, and the variation in the oscillator frequency is compensated. Thus, the oscillator frequency remains equal to a whole multiple of the frequency of the impulses applied to the first grid.

If this synchronism has not occurred at the moment of switching on, there will be a periodic variation of the voltage produced across C_3 , in other words an alternating voltage is obtained⁷⁾ by which, if the frequency is lower than the cut-off frequency of F , the oscillator voltage is frequency modulated. If, during this modulation, the oscillator frequency passes a value which is equal to a whole multiple of the impulse frequency, the oscillator frequency will remain at this value. Theoretically, it should be of no account whether the relative phasing of V_{g1} and V_{g3} adjusts itself in such a way that the impulse lies upon one or the other flank of the sinusoidal curve. According to the direction of the control of B_2 , the adjustment upon one of the flanks is, however, unstable and that upon the other is stable.

The low-pass filter F suppresses voltages at the impulse frequency or multiples of it. This prevents the oscillator from being subjected to phase modulation at these frequencies.

When compiling fig. 6, the oscillator frequency was taken to be equal to twice the impulse frequency. It will be clear that synchronization, as described above, may also occur when the ratio between these numbers is a larger integer. The maximum value of this number, as will be obvious from fig. 6, is dependent upon the duration of the impulses. For, if the frequency of the oscillator voltage is so high that the duration of one period is equal to the duration of the impulses, the time integral of the anode current impulses will be independent of the mutual phase position of V_{g1} and V_{g3} , so that, with a phase shift of the oscillator voltage with respect to the voltage impulses, no alteration occurs in the control voltage⁸⁾. For still higher frequencies, a control effect does again occur, reaching a maximum when the ratio of the impulse duration to the duration

⁷⁾ This effect is analogous to the stroboscopic effect in light which occurs, for example, when a rotating wheel with spokes is illuminated by light flashes, whereby the "spoke frequency" deviates a little from a whole multiple of the flash frequency. The wheel is then seen to rotate slowly either clockwise or anti-clockwise.

⁸⁾ A similar problem is described by J. M. L. Janssen, Philips techn. Rev. 12, 52-59, 1950 (No. 2), in particular on pages 57 and 58,

of one period of the oscillator voltage reaches a value of $3/2$. When the ratio is 2, there is again no control effect, etc. Fig. 7 illustrates, on a relative scale, the control voltage as a function of the above-mentioned ratio. Here, the control voltage is the variation of the voltage across the capacitor C_3 at a change in phase of 180° of V_{g3} with respect to V_{g1} . In order to ensure the correct operation of an IGO, the duration of the impulses should preferably

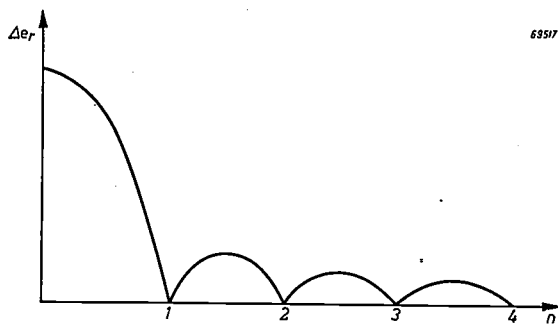


Fig. 7. Variation Δe_r of the control voltage at a phase change of 180° of the oscillator voltage as a function of the ratio n of the duration of the impulse to that of a cycle of the oscillator voltage.

not be larger than half a period of the oscillator voltage. To facilitate the synchronization of an oscillator whose frequency is much higher than the impulse frequency, it will therefore be necessary to make the impulse duration as short as possible. In practice, this is limited by the fact that, at very short impulses, the average value of the current through the valve B_3 (see fig. 5) becomes extremely small. Usually, however, another practical limitation occurs at even lower oscillator frequencies. If, for example, it is required to synchronize the oscillator voltage with the 200th harmonic of the impulse voltage, this means that a deviation of only $1/2\%$ in the proper frequency of the oscillator gives a neighbouring harmonic of the impulse voltage. Hence, high requirements are already imposed on the frequency stabilization of the controlled oscillator itself, in order to maintain the synchronization at the required frequency. For this reason, it appears to be impracticable to make Impulse Governed Oscillators which are synchronized at higher harmonics of the pulse voltage than the 200th.

In an IGO, a distinction can again be made between a collecting and a holding zone, with regard to the frequency range in which the frequency stabilization is operative. The former frequency range is determined mainly by the cut-off frequency of the filter F (fig. 5). For, before the synchronization of the oscillator voltage occurs, the mixer valve B_3 supplies an alternating voltage and only when the frequency of this voltage is so small that it is passed

by the filter F , is it possible for frequency stabilization to operate. The "holding" zone is determined by the control voltage, supplied by B_3 , when, once synchronization has occurred, the oscillator voltage is subjected to a certain phase shift. Moreover, it depends on the frequency range in which the oscillator can be detuned by the reactance valve. As there is a possibility that the oscillator voltage is synchronized to a wrong harmonic of the impulse voltage, it is desirable not to make the latter frequency range much larger than the frequency interval between two harmonics. By making provision that the holding zone slightly exceeds this interval, it may be achieved that with the (apparently) continuous detuning of the oscillator, the frequency jumps steadily from one harmonic to the next. The construction can then be made in such a way that, by means of a pawl setting, the correct harmonic is selected.

For most applications oscillators must supply a pure sinusoidal voltage with the least possible amplitude and frequency modulation. Deviations from the sine wave may be mainly due to two causes, in an IGO. These are:

- the filter F (see fig. 5) does not sufficiently suppress voltages with the impulse frequency and, possibly, harmonics of it; this causes frequency modulation of the oscillator;
- harmonics of the impulse voltage produce an alternating voltage in the oscillator circuit via the mixer valve B_3 (see fig. 5); when a multigrind mixer valve is used, this can occur only via the interelectrode capacitances.

Practice has shown that in the case of frequencies which are not too high, it is comparatively easy to ensure that the voltages with unwanted frequencies are at a level of about -80 dB with respect to the oscillator voltage, mainly because the power of the impulse harmonics is small compared with the output of the oscillator. This level is practically always sufficiently low.

It has been shown that an IGO is a circuit fed with a periodic series of voltage impulses, and that it supplies an alternating voltage of which the frequency is equal to that of one of the harmonics of the voltage impulses. The IGO is, therefore, to a certain extent similar to a filter which passes such a narrow frequency band that only the required harmonic is passed. The frequency to which this fictitious filter has been adjusted may be changed simply by varying the tuning of the oscillator, whereas the width of the frequency band is determined by the

low-pass filter F . For instance, when an interfering alternating voltage is present on the first grid of the valve B_3 , in addition to the voltage impulses, and the frequency of this interfering voltage differs very little from that of the required harmonic of the impulse frequency, an alternating voltage will originate in B_3 of which the frequency is equal to the difference of the frequencies of the oscillator voltage and the interfering voltage. It depends entirely upon the filter F , whether this voltage is passed to the grid of the reactance valve or not. If it is, the oscillator voltage will be frequency modulated. If the filter F does not pass the interfering voltage, the oscillator will not be affected by it. It is of importance, in this respect, to choose the cut-off frequency of the filter F as low as possible (unless it is desired to modulate the oscillator frequency). In this case the collecting zone will become small and, in practice, a compromise will have to be accepted.

A circuit like the one depicted in fig. 5 is a form of inverse feedback. If a phase shift occurs in the circuit through which the control voltage of the mixer valve B_3 is supplied to the grid of the reactance valve, there is a possibility that the circuit becomes unstable in this sense that the oscillator does not operate at a constant frequency, but is subjected to periodic frequency changes, thus is frequency modulated. This phenomenon is almost identical to what happens in other circuits with inverse feedback. There is, however, one fundamental difference in the stability condition. An alternating voltage applied to the grid of the reactance valve B_2 gives rise to frequency modulation, whereas the control voltage supplied by the frequency changer B_3 varies through phase shifts of the oscillator voltage. There is a phase shift of 90° ⁹⁾ between a phase change (displacement) and frequency change that corresponds with it. Whereas in an "ordinary" amplifier with inverse feedback, a phase shift of 180° is necessary at the critical frequency¹⁰⁾ to cause instability, here it is merely 90° . In addition, the phase change (displacement) which belongs to a certain frequency change is inversely proportional to the modulation frequency. This means that the higher the critical frequency can be adjusted, the smaller will be the risk of instability in an IGO.

In the example so far reviewed it is assumed that an oscillator is synchronized by an impulse generator. It is, however, also possible to synchronize an impulse generator with an oscillator of which the frequency has been stabilized by a crystal. This may be done by feeding the control voltage supplied by the mixer valve to a control system which influences the frequency of the impulse oscillator. If the frequency of the impulse generator is lower than that of the crystal oscillator, frequency division will take place in which the frequency ratio is the same as the multiplication factor in the previous case.

When an impulse generator is phase modulated, an IGO synchronized to the n th harmonic of it will also be subjected to an n times larger phase modulation, provided the time constant of the control circuit is small enough to pass the highest modulation frequency without much phase shift. Because of the high value of n , which can be reached quite easily in this case, an extensive phase shift can be obtained without much trouble. Thus, a "phase modulation in a narrower sense"⁹⁾ is obtained. In order to get the form of the phase modulation which is usually referred to as frequency modulation, the modulation voltage must be supplied via a network which will correct the frequency characteristic in this sense that the "phase swing" becomes inversely proportional to the frequency of the modulating voltage.

Practical construction of the impulse generator

For most purposes in which an IGO is used, the impulse frequency must comply with high demands of accuracy and stability. The frequency must have a very accurate value, for example, exactly 100 kc/s, so that the IGO is always synchronized to a frequency which is a whole multiple of this value. In view of the fact that, in practice, synchronization to a frequency which is 200 times higher than that of the impulse generator is possible, an IGO may be constructed with a frequency up to 20 Mc/s. As it is preferable for the impulse duration not to be longer than the duration of a half cycle of the voltage produced by the IGO, the maximum impulse time is about $1/40$ μ sec.

Because of the considerable stability required, an impulse generator is always used, in practice, in which the impulses supplied are synchronized, in their turn, with a sinusoidal voltage. The latter is supplied by an oscillator of which the frequency is stabilized by a quartz crystal, which may be placed in a thermostatically controlled container. Fig. 8 gives details of a circuit which might be used for this purpose. This is based upon the principle of the ringing oscillator. The hexode section of the triode-hexode B_1 fulfils this function. By a suitable selection of the grid capacitor C_1 and the grid leak R_1 , intermittent oscillation is obtained¹¹⁾. The setting may be adjusted so that the oscillating is interrupted after a few periods and does not start again until the capacitor C_1 , which is charged because of the flowing of a grid current, has been sufficiently discharged via the resistor R_1 . Thus, a series of periodic, damped, oscillations occurs in the anode circuit of the hexode section, as is indicated in the lower part of fig. 8. Its repetition frequency is largely determined by the values of C_1 and R_1 .

⁹⁾ See the article mentioned under ²⁾, p. 44.

¹⁰⁾ See B. D. H. Tellegen, Inverse feed-back, Philips techn. Rev. 2, 289-294, 1937.

¹¹⁾ See J. van Slooten, The functioning of triode oscillators with grid condenser and grid resistance, Philips techn. Rev. 7, 40-45, 1942, and Stability and instability in triode oscillators, Philips techn. Rev. 7, 171-177, 1942.

The triode section of B_1 is used as an oscillator of which the frequency is stabilized by a quartz crystal. The oscillator voltage is applied to the third grid of the hexode section. When the repetition frequency of the damped oscillations is approximately the same as the frequency of the crystal oscillator, synchronization occurs, and these frequencies will become identical. The damped oscillations obtained

been constructed in accordance with the principles described here. The chassis also houses the mixer valve (B_3 in fig. 5). The impulse duration which can be obtained with this device is $1/14 \mu\text{sec}$. Proceeding from the condition that the duration of the impulses may, as a maximum, be equal to the duration of a half cycle of the voltage generated by the IGO, it will be clear that this may be used

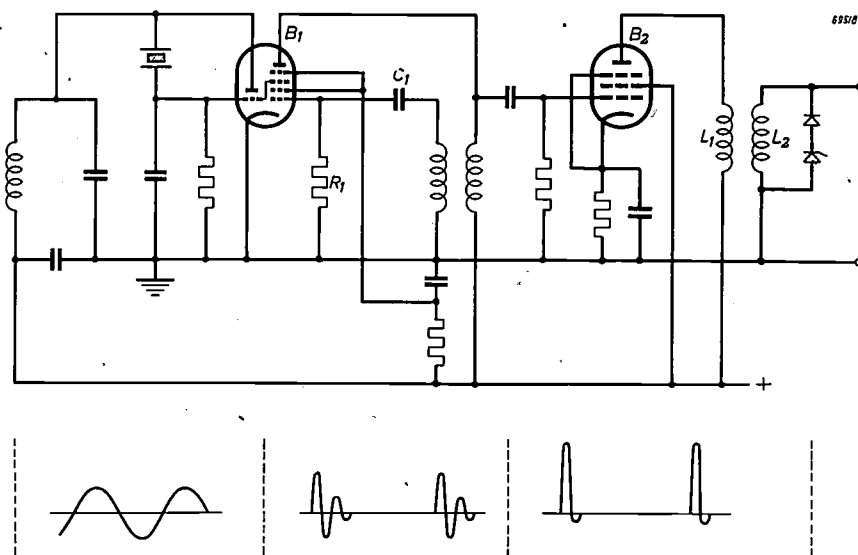


Fig. 8. Diagram of an impulse generator. The triode section of the valve B_1 acts as a crystal-controlled oscillator; the hexode section of this valve forms a ringing oscillator. The impulse duration is shortened by means of the valve B_2 . The lower part of the illustration shows the anode currents as functions of the time.

are now applied to the grid of the valve B_2 , which is negatively biased to such an extent that anode current flows only during part of the first half cycle of each of the series of damped oscillations. (For, the amplitude is largest in the first half cycle). Thus, short current impulses occur in the anode circuit of B_2 . The coil L_1 has been incorporated in this circuit and with the valve and wiring capacitances it forms an oscillatory circuit. This circuit is periodically excited by the anode current impulses. A second coil, L_2 , to which a rectifier has been connected, has been coupled to the first coil L_1 , so that this oscillatory circuit can pass on only half an oscillation every time it is excited, thus ensuring voltage impulses of very short duration. These pulses serve to synchronize the actual, adjustable oscillator (see fig. 5).

Synchronization of the intermittently oscillating oscillator with the crystal oscillator may also be done so that frequency division takes place. The crystal oscillator may oscillate, for example, at a frequency of 100 kc/s, while the periodic oscillating occurs at a frequency of 50 kc/s.

Fig. 9 illustrates an impulse generator which has

to synchronize an oscillator with a maximum frequency of 7 Mc/s. By adding a third valve to a set-up as shown in fig. 8, connected in the same way as B_2 , it has shown possible, in practice, to obtain an impulse duration of $1/180 \mu\text{sec}$, so that the synchronization of an oscillator with a frequency of 90 Mc/s is possible.

Decade tuning

It has been explained that an IGO can be synchronized to a frequency which is equal to any multiple of the frequency of an impulse generator, which itself can be synchronized by a crystal oscillator. This makes the IGO admirably suited for use in transmitters, receivers and other apparatus in which so-called decade tuning is used. The principle of this system has already been explained with the aid of fig. 1. An alternating voltage with a frequency, which can be varied over an extensive range, is obtained by supplying two voltages to a mixer valve. One of these voltages is adjustable to frequencies which are a whole multiple of, for example, 100 kc/s. The frequency of the other oscillator is continuously variable in a range of 100 kc/s, for

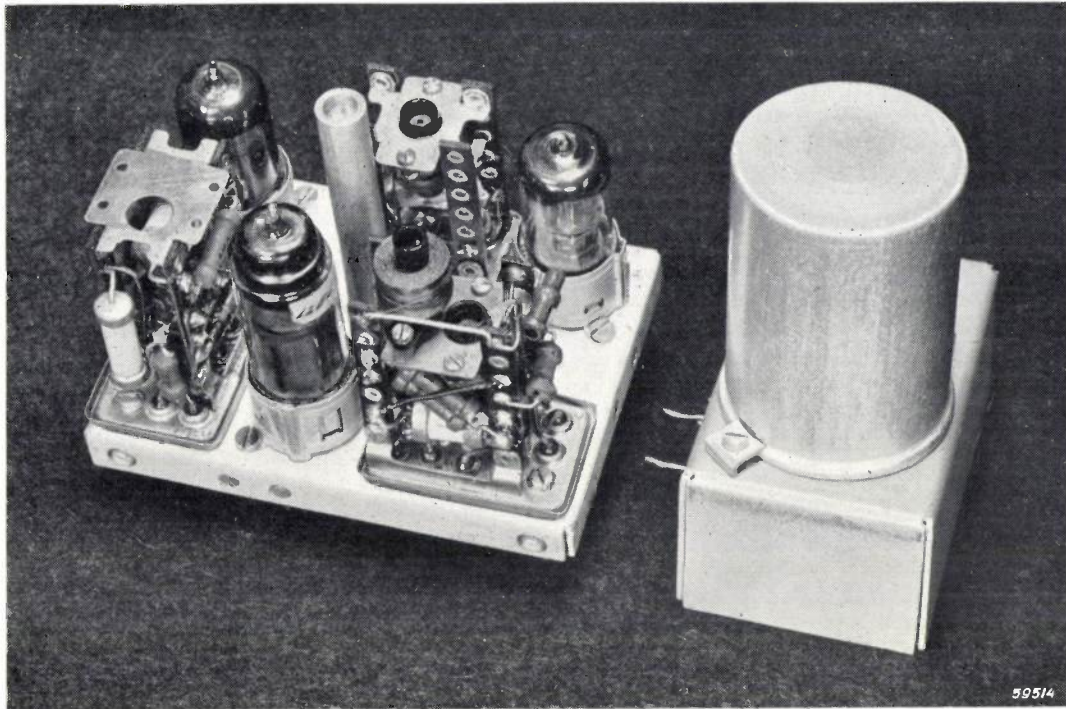
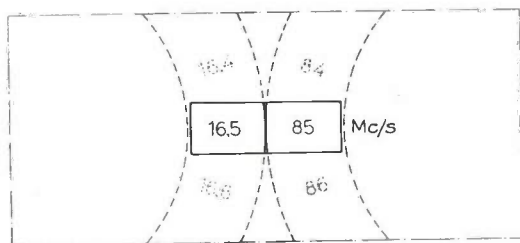


Fig. 9. Crystal-controlled impulse generator and impulse mixer valve (fig. 11: *IG* and *IM*) from the aircraft transmitter SVZ 101.

example, from 200 kc/s to 300 kc/s. The mixer supplies a voltage with a frequency which is equal to the sum of the frequencies of the two oscillators. The stages of 100 kc/s are adjusted by means of a single tuning device, for example, by means of a mechanical pawl system, and a second adjustment device ensures the fine control within the range of 100 kc/s. By placing the two dial controls next to each other, a direct-reading dial is obtained which combines the figures of both scales into one number (fig. 10). The accuracy of the readings will be the same for all frequency ranges.

The accuracy with which the required frequency can be adjusted depends upon the frequency accuracy of the two oscillators. The oscillator which is continuously adjustable in the range from 200 to 300 kc/s can be very exact because of the low frequency and the relatively small range. It is much



73649

Fig. 10. If the two dials, in decade tuning, are mounted next to each other, a direct-reading scale is obtained on which the figures of both dials are combined (in this case 16585 Mc/s).

more difficult, however, to construct an oscillator which can be tuned in steps of 100 kc/s and is sufficiently exact at each one of these stages. It will be clear from the above that the IGO is a most attractive means of solving this problem. By synchronizing the impulse generator used with an oscillator of which the frequency has been fixed at 100 kc/s by means of a crystal, the controlled oscillator can be synchronized to every required multiple of this figure and with the same degree of accuracy as that of the crystal-controlled oscillator.

The IGO in an aircraft transmitter

Fig. 11 depicts the fundamental diagram of an aircraft transmitter (Philips Type SVZ 101) which incorporates an IGO. The frequency range of this transmitter is from 2.8 to 24 Mc/s. Twelve frequency bands, each having a width of 100 kc/s, can be selected by means of pawl control knobs¹²⁾.

The impulse generator *IG* has been constructed according to the principle shown in fig. 8. The crystal-controlled oscillator operates at a frequency of 100 kc/s and produces impulses with a frequency of 50 kc/s. The impulse generator and impulse mixer valve *IM* have been constructed as one unit, as illustrated in fig. 9. The IGO which can be adjusted

¹²⁾ A very suitable pawl control knob system for this purpose was described in: W. L. Vervest, Automatic tuning of transmitters, *Communication News*, 10, 20-29, 1949.

between 1.55 and 6.1 Mc/s will, therefore, always be synchronized to a frequency which is a whole multiple of 50 kc/s. The alternating voltage supplied by the IGO is used to synchronize the driver stage *ST*. The alternating voltages of the IGO and the driver stage are supplied to the mixer *M* to this end. This supplies, inter alia, an alternating voltage with a frequency equal to the difference between the frequencies of the driver stage and the second harmonic of the IGO. Together with the alternating voltage produced by an oscillator *VO* with continuously variable frequency, the alternating voltage

IGO ranges from 1.55 to 6.1 Mc/s, and that of the variable oscillator from 0.2 to 0.3 Mc/s, the driver stage may be adjusted to frequencies between $2 \times 1.55 - 0.3 = 2.8$ Mc/s and $2 \times 6.1 - 0.3 = 12$ Mc/s. The output stage *ET* is used as a frequency doubler if higher frequencies are required, so that the highest transmitting frequency amounts to 24 Mc/s.

The variable oscillator *VO* is housed in a small container which is mounted some distance away from the transmitter. *Fig. 12* shows this container. Because of the low frequency and the small fre-

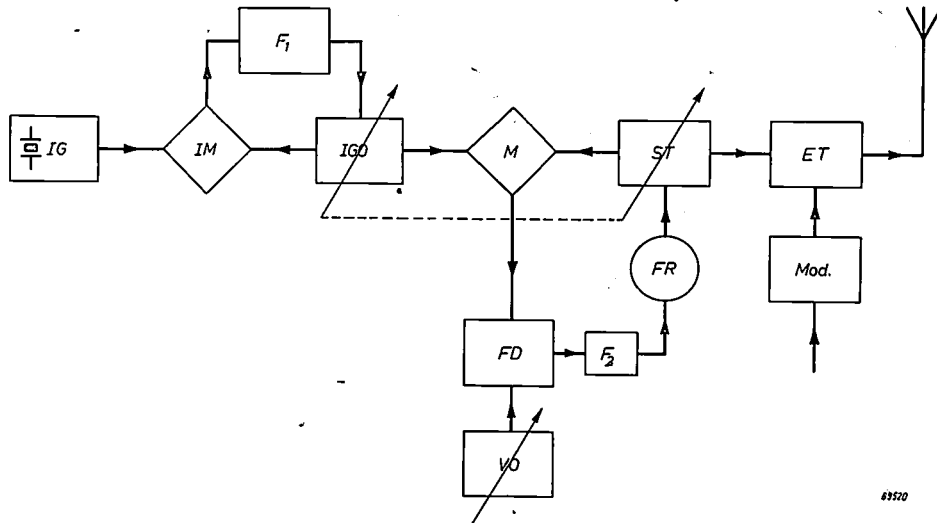


Fig. 11. Fundamental diagram of the Philips aircraft transmitter SVZ 101. *IG* = impulse generator, *IM* = impulse mixer, *IGO* = impulse-governed oscillator, *F₁* = low-pass filter, *M* = mixer, *ST* = driver stage, *FD* = phase discriminator, *VO* = oscillator with continuously variable frequency, *F₂* = low-pass filter, *FR* = frequency control device, *ET* = output stage, *Mod.* = modulation amplifier.

supplied by *M* is applied to the phase discriminator *FD*, which supplies a control voltage which is fed, via the filter *F₂*, to the frequency control device *FR*, which varies the frequency of the driver stage *ST*. The last-mentioned frequency will always adjust itself to a value which is equal to the difference between the second harmonic of the IGO and the frequency of the variable oscillator *VO*. The last-mentioned is variable between 200 and 300 kc/s, so that the frequency of the driver stage can be varied, by means of *VO*, over a frequency band of 100 kc/s. The driver stage *ST* and the IGO may be adjusted by means of one control knob to the required frequency band, and a padding capacitor maintains the frequency difference between *ST* and the second harmonic of the IGO to about 250 kc/s. Great accuracy is not necessary here, for, as has already been explained, the frequency of the driver stage is exactly determined by that of the IGO and the variable oscillator. As the frequency range of the

frequency range, the frequency stability of this oscillator is very high. In addition, various other measures have been taken to increase this stability as much as possible. The oscillator valve, for instance, has been coupled to the circuit as loosely as possible and special capacitors have been incorporated to compensate for variations in capacitances due to temperature changes. All elements which are part of the oscillatory circuit, i.e. the coil, the variable capacitor and the trimmers, are contained in a box made of a material which is a bad heat conductor, so that the temperature of these parts will be substantially the same. Because of these precautions, the frequency variation due to a temperature change of 60 °C was limited to 150 c/s.

The frequency control device *FR* influences the frequency of the driver stage *ST* in two ways, namely, both by means of a reactance valve, shunted across the oscillator circuit and by a regulating coil, also shunted across this circuit, and wound

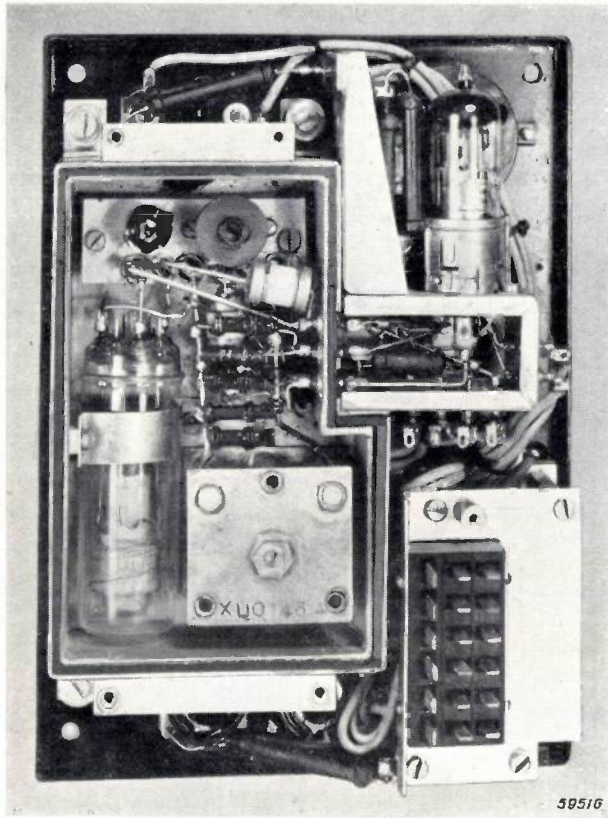


Fig. 12. Control box of the aircraft transmitter SVZ 101. It contains an oscillator with continuously variable frequency between 200 and 300 kc/s.

on a Ferroxcube core. The permeability of this material is varied by the magnetic field of a yoke which is energized by direct current which flows through a coil on the yoke. As the last-mentioned coil has a fairly high inductance, this control is slow, but the frequency variation that can be obtained in this manner is much larger than that which is attained with a reactance valve. On the other hand, control by means of a reactance valve enables compensation of very rapid frequency variations.

Further particulars of the aircraft transmitter SVZ 101 will be found elsewhere ¹³⁾.

Other applications of Impulse Governed Oscillators

Finally, there are some other branches of telecommunications in which an IGO might be useful.

In the first place this system offers many possibilities for generating carriers for carrier telephony. In general, these carrier frequencies are multiples of a certain basic frequency, for example 4 kc/s. Each of these frequencies can be generated by an IGO, and these oscillators are all synchro-

nized by means of one impulse generator. This may lead to considerable simplification compared with a system whereby the different carriers are obtained by means of filters from an impulse generator with a frequency of 4 kc/s ¹⁴⁾.

One or more impulse-governed oscillators may, with advantage, be used for generators used for frequency standardization, because it is possible to obtain, from one reference frequency, any frequency with great accuracy. Because of decade tuning, the frequency can be read directly, which practically eliminates errors, interpolations, so often used in other systems, not being necessary here.

By applying the IGO principle in receivers which are constructed according to the so-called double superheterodyne principle, the dial may also be constructed according to the decade system, so that the accuracy of adjustment will be almost the same for all frequencies. The first frequency transformation might be made by an IGO, making frequency steps of 100 kc/s, and the second by means of an oscillator of which the frequency is continuously variable in a range of 100 kc/s.

In transmitters and receivers which have been arranged according to the IGO principle, any frequency up to 30 Mc/s can be found immediately: i.e. there is no "searching". This ensures faster and more reliable traffic. It also makes possible, because of the great choice from very many channels which are close together in frequency (for example a distance between them of only 1 kc/s for telegraphy), the dividing of the traffic over these channels, thus lessening the risk of interference and reducing the inevitable waiting time. This is of utmost importance for aircraft and navigation services.

¹⁴⁾ See: D. Goedhart and G. Hepp, Carrier supply in an installation for carrier telephony, Philips techn. Rev. 8, 137-146, 1946.

Summary. This article describes a method of constructing an oscillator which can operate at many different frequencies, yet each possesses the same stability as the frequency produced by a crystal oscillator. Only one crystal is necessary. To do this, the oscillator is synchronized with a harmonic of a periodic impulse (Impulse-Governed Oscillator, IGO), which is obtained from an impulse generator whose frequency is determined by a crystal. The system is similar to a filter which passes a very narrow frequency band and whose tuning frequency can be changed by means of a single circuit.

An IGO permits frequency multiplication and division in a ratio of about 1:200. A combination of an IGO and an oscillator with a continuously variable frequency which is much lower than the frequency of the IGO and which, therefore, possesses a high degree of stability, can be used in a transmitter, which can operate at any required frequency in a certain range and possesses great frequency stability. Other branches of the telecommunication technique, in which an IGO can be used to advantage, are: carrier telephony, frequency standards and radio receivers.

¹³⁾ E. H. Hugenholtz, The application of impulse-governed oscillators (IGO) in aircraft transmitters, Communication News 11, 13-21, 1950 (No. 1).

COAXIAL CABLE AS A TRANSMISSION MEDIUM FOR CARRIER TELEPHONY

by H. N. HANSEN *) and H. FEINER *).

621.395.44:621.315.212

For many years carrier telephone systems have been operated over two different types of cable, namely quadded and coaxial. During this period of rival development, both techniques have gradually improved and both have been successfully adopted by Telephone Administrations in various countries to build extensions of the European long-distance cable network.

This article reviews the history of the development period mentioned above and discusses the comparative advantages of the two systems in a qualitative way. It finally expresses the opinion that neither system has yet achieved a decisive advantage over the other in all fields of application.

Introduction

The justification for carrier telephone systems lies in the economic advantage of providing several speech channels over each physical telephone circuit formed by a pair of line conductors. Before carrier working was introduced, only one audio channel was obtained from each physical circuit, with the result that long-distance telephone systems were composed of large multi-conductor cables whose cost formed a high proportion of that of the entire system. It was, therefore, advantageous to reduce the cable cost by increasing the number of channels per pair of conductors.

At first, in the early 1920's, carrier systems were operated on open-wire lines and, somewhat later, on a few cable circuits. By about 1930 carrier telephone systems with 12 channels were introduced, one such system being operated over each pair of conductors of a paired or (somewhat later) "quadded" cable, the latter containing groups of four conductors arranged in a star. At about the same time, in the United States experiments were being carried out with a type of cable which would admit of several hundred channels per pair, viz. the coaxial cable, containing pairs of concentric conductors. The first experimental multi-channel coaxial carrier telephone system was installed in 1929 at Phoenixville, Penn., by the American Telephone and Telegraph Corp.¹⁾ About 20 000 miles of coaxial cable are now in operation in the U.S.A.; the transmission system is designed to provide a maximum of 600 channels per coaxial tube. Although the above is only a fraction of the total amount of telephone cable in use, the coaxial cable has clearly proved its worth in a very short time.

Both the above systems were also adopted in Europe. Between 1930 and 1940, coaxial carrier telephone systems were introduced in Great Britain, Germany and France²⁾ to work alongside the quad cables, on which 12-channel systems were installed. During and after the second world war, the Netherlands Post, Telephone and Telegraph Service were able to show that on carefully constructed quad cables, the number of channels per pair could be considerably increased, namely from 12 or 24 up to 48³⁾. This discovery had some reactions on the development of the European cable network, but interest in coaxial systems was nevertheless maintained, and steadily increased. In Great Britain a very considerable number of coaxial systems has already been established, and the P.T.T. in France is also engaged on similar activity.

It seems to us to be opportune to present in this Review a brief survey of the rival development of coaxial and quad cables, and to attempt to assess qualitatively their relative advantages and expected fields of application. This will also provide a convenient introduction to future publications describing the work of our laboratories in the coaxial cable field. Two such papers have already appeared⁴⁾.

*) N.V. Philips Telecommunication Industry, Hilversum.

¹⁾ Bell Lab. Record 27, 234, 1949 (Coaxial cable's 20th anniversary). See also L. Espenschied and M. E. Strieby, Systems for wide-band transmission over coaxial lines, Bell. Syst. Techn. J. 13, 654-679, 1934.

²⁾ Siemens Veröff. a.d. Geb. d. Nachr. 6, issue of 28th Jan. 1937 (series of articles).

A. H. Mumford, The London-Birmingham coaxial cable system, The Post Office Electrical Engineers' Journal 30, 206, 270, 1938; 31, 51, 132, 1938.

R. Sueur, L'évolution de la technique des lignes a grande distance depuis 15 ans, Ann. Télécomm. 6, 146-164, 1951.

³⁾ G. H. Bast, D. Goedhart and J. F. Schouten, A 48-channel carrier telephone system, Philips techn. Rev. 9, 161-170, 1947.

⁴⁾ J. F. Klinkhamer, A through supergroup filter for carrier telephone systems on coaxial cable, Philips techn. Rev. 13, 223-235, 1952. (No. 8).

H. N. Hansen A new supergroup modulation scheme for coaxial cable telephone systems, Comm. News. 12, 1-9, 1951 (No. 1).

General description of a carrier telephone system

A carrier telephone system may be classified in three categories, as shown in *fig. 1*, viz.:

- the carrier terminal equipment,
- the cable,
- the line equipment.

The carrier terminal equipment for each physical circuit comprises, at the sending end, a modulation system which enables each channel from the audio-frequency range to be translated upwards in frequency to some specified location in the frequency spectrum. At the receiving end, analogous demodulation equipment translates the incoming high-frequency channels downwards to their original audio-frequency range. The nominal width of each channel is usually 4 kc/s. We shall not deal extensively with the actual modulation process. Usually

is associated with an equalising network, so designed that the net overall gain frequency curve of each repeater (by which we mean a line amplifier combined with its equaliser) will match the loss-frequency curve of a repeater section of cable rather accurately. In this way, cumulative linear distortion is avoided, in all channels.

The number of channels per pair

The above-mentioned increase in the number of channels per pair entails a corresponding increase in the frequency bandwidth which must be transmitted over each physical circuit. On cables of the types under discussion, the attenuation increases with frequency. An increase in bandwidth thus leads to an increase in the total amount of gain required throughout the circuit.

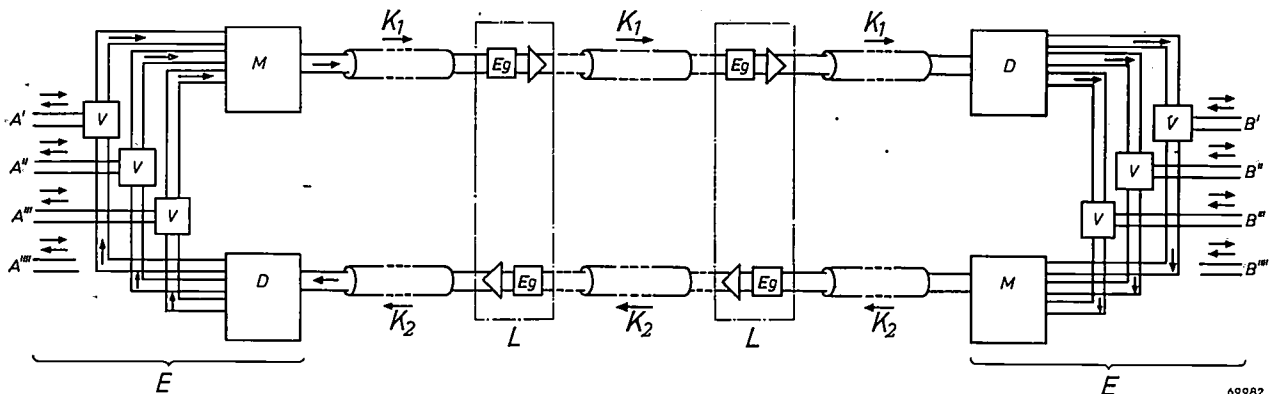


Fig. 1. Block schematic of a carrier telephone system. E = carrier terminal equipment to which the subscriber's lines $A', A'', \dots B', B'', \dots$ are connected through 4-wire terminating sets, V . The outgoing signals are passed to the modulating apparatus M and the incoming signals to the almost identical demodulating apparatus D . K_1, K_2 are the "go" and "return" cables. L = line equipment installed in repeater stations. In each station an equaliser Eg and amplifier are provided for each cable pair in each direction.

69982

modulation is done in several steps: By a first transposition to higher frequencies, 12 audio-frequency channels are assembled to form a group; 5 such groups (or less) by a second transposition are combined to form a supergroup. If there are more than 60 channels, each of the supergroups again must be shifted to its final frequency range (the supergroups together form a "hypergroup").

In quad cable systems, up to the present time it has been normal practice to provide separate "go" and "return" cables for the two directions of transmission.

Line amplifiers must be installed at repeater stations spaced at intervals (repeater sections) along the route; the amplifier gain compensates for the attenuation to which the signals are subjected during their propagation along the cable. In each station, the line amplifier for each physical circuit

This will necessitate a shorter repeater spacing, as can be readily understood. The signal arriving at a repeater must not be attenuated by the preceding cable section to a degree that it will suffer from the thermal noise of the cable and the first amplifying valve of the repeater. Hence, there is a minimum permissible input level. On the other hand, the output level is limited by the power available from the power valves. Thus, the gain to be provided by one repeater is fixed, and a higher total gain can only be obtained by increasing the number of repeaters.

Nevertheless, an increase in the number of channels will economically be justified by the saving in cable cost per channel. Moreover, the number of repeaters per channel actually will also be reduced by virtue of the fact that the increase in cable attenuation (i.e. the required gain) is less than proportional to the frequency (it will usually be

found to increase as the square root of the frequency).

There are, however, several limitations which prevent an indefinite increase in bandwidth. Since thermal noise is introduced at the input of every repeater, the minimum permissible level of the signals at the end is dependent on the number of repeaters in circuit. Similarly, the total output power available from a repeater diminishes with increasing frequency; moreover, this power has to be shared between a greater number of channels, so that the maximum attainable level of the signals at the output end is reduced when the bandwidth is increased. The maximum gain obtainable is seriously affected by these effects: for example in a 48-channel system for use on quad cable, the maximum amplifier gain is about 65 dB, but in a 960-channel coaxial system it is not more than 50 dB. Consequently, the repeater spacing must be proportionately shortened to allow for these level limitations, and it is fairly clear that there must be some theoretical optimum beyond which it is not economical to increase the bandwidth.

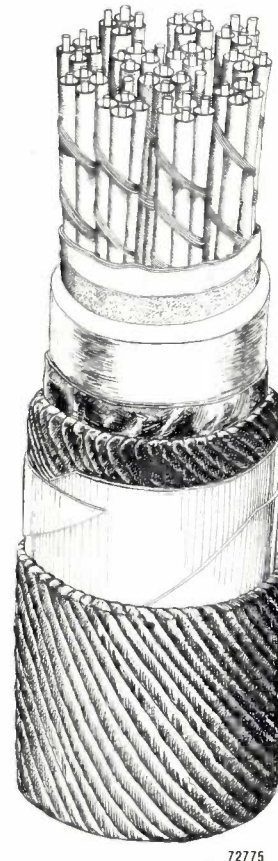
In systems in use today, operating on either quad or coaxial cables, this economic limit has by no means been reached. The practical limit to the number of channels per pair is imposed by characteristics of the components at present available, in coaxial systems particularly the valves. This limit cannot be raised by any reasonable expedient; for example, it is useless to increase the number of valves per amplifier. These limitations will be explained presently.

Quadded telephone cable

Consider first the common quadded cable, containing pairs of conductors, every two pairs being combined into stars (star quads). Fig. 2 shows a 12-quad cable of the type widely employed by the Netherlands P.T.T. Air-spaced paper insulation is adopted, and the cable is covered by a lead sheath as a protection and screen, and to prevent the ingress of moisture. Further protective layers are provided to prevent physical damage to the cable if it is laid directly in the ground.

The cable attenuation is shown as a function of frequency in fig. 3. The absolute value of the attenuation, other things being equal of course, depends on the diameter of the conductors. The choice of conductor diameter, within limits imposed by electrical and mechanical requirements, is governed by economic considerations again; large diameters result in costly cables, but reduced attenuation and, therefore, fewer repeaters. The most economic choice will be greatly affected by the market prices of

copper and lead. In practice it is found that the optimum conductor diameter lies between 0.9 and 1.3 mm. The attenuation at 200 kc/s is about



72775

Fig. 2. Typical 12-quad cable. Each star quad contains two diagonal pairs. Air-spaced paper insulation is used. The cable is sheathed with lead. Outer servings and steel tape armouring are applied over the sheath.

2.4 dB/km for 1.3 mm conductors. Assuming a repeater gain of 60 dB, the repeater spacing for such a cable is 25 km, which corresponds with normal practice applicable to the 48-channel system used

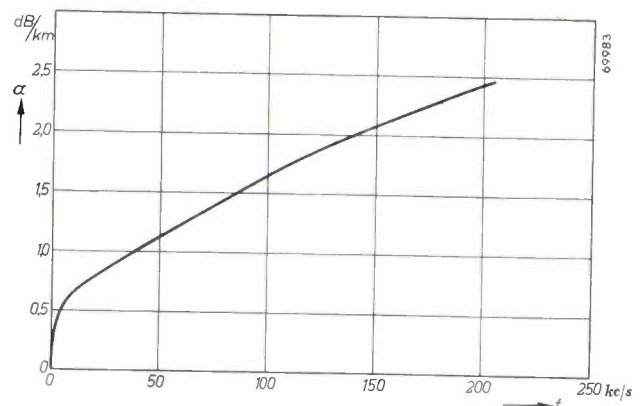


Fig. 3. Attenuation-frequency characteristic of a 1.3 mm pair in a quad cable. The diagram shows the attenuation α in dB/km as a function of frequency f (in kc/s).

in the Netherlands and elsewhere (frequency band 12-204 kc/s, repeater spacing about 26 km).

So far, it has not been possible to increase considerably the number of channels per pair in quad cable due to limitations of crosstalk. Crosstalk — the transfer of signals between different pairs of conductors — is due to the effects of capacitance unbalance and mutual impedance between the various pairs in the cable. Crosstalk in quad cables increases with the frequency of the signals. By suitable precautions in the manufacture and installation of modern quad cables⁵⁾, it has been found possible to keep the crosstalk within permissible limits up to a frequency of 250-300 kc/s, which is sufficient to allow the operation of 60-72 channels per pair. Possibly some further progress may yet be made, but the most optimistic estimate of what may be achieved in the future does not exceed 120 channels per pair.

If such an increase in the number of channels per pair is ever to be attained, a very efficient balancing technique will have to be evolved. For this purpose, the introduction of variable balancing elements at intermediate points within each repeater section is contemplated.

Attention today is concentrated on possible ways of diminishing the crosstalk because new methods of cable construction promise an increase of frequency bandwidth without a corresponding reduction in the repeater spacing. The attenuation of present types of cable is partly attributable to dielectric losses in the paper insulation, especially at frequencies in the order of 200 kc/s, the loss angle $\tan \delta$ then reaching a value of 120×10^{-4} . Various manufacturers are developing new types of cable in which the insulating material consists of polyethylene, styroflex or polystyrene foam, whose low dielectric constant and loss angle ($\tan \delta = 3 \times 10^{-4}$ or less) enable the HF attenuation to be reduced considerably below the figures applicable to paper cable. In fact, the attenuation of foam-insulated cables at 550 kc/s is comparable with that of paper-insulated cables at 250 kc/s.

Coaxial telephone cable

A coaxial pair or tube generally consists of a copper wire which forms the inner conductor or core, supported by some insulating structure, so that its axis coincides with that of a tubular outer conductor. The outer conductor is usually also of copper, but is sometimes made of aluminium, which is cheaper. Several such pairs or tubes may be laid up together with auxiliary conductors to form a coaxial cable. The auxiliary conductors are required to provide a service telephone, extended alarm circuits etc. Fig. 4 shows an 8-tube coaxial cable; a type has been standardised in the United States. The diameters D and d of the outer and inner conductor

respectively in these cables, both being made of copper, (fig. 5), bear roughly the ratio:

$$\frac{D}{d} = 3.6 \dots \dots \dots (1)$$

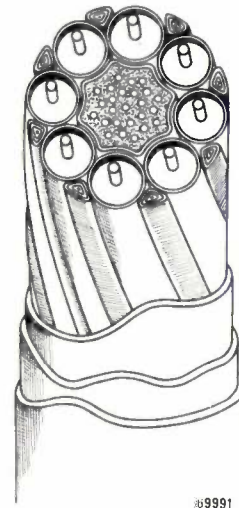


Fig. 4. Typical coaxial cable as used in the USA. Eight coaxial tubes are laid up helically over a core consisting of auxiliary signal wires. Similar sheathing and armouring as fig. 2.

In one cable standardised internationally, $d = 2.6$ mm and $D = 9.4$ mm ($3/8''$). The inner conductor is supported along the axis of the tube by discs or beads of polythene or other materials, spaced 1'' apart, or alternatively by helical band of polythene or styroflex, etc. The remaining space is filled with dry air or gas, sometimes under pressure.

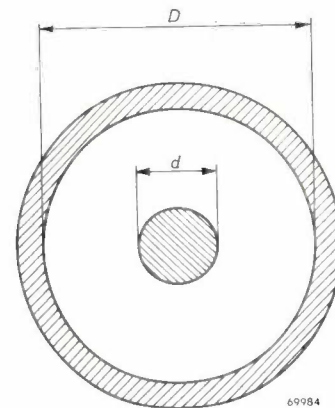


Fig. 5. Cross-section of a coaxial tube. The ratio of diameters D/d is usually 3.6.

The motivation of the condition (1), as well as a brief summary of the transmission characteristics of coaxial cables, will be found in the Appendix to this article.

We can now state in a few words the reason why it has been possible to employ such a very much greater number of channels per pair in the coaxial cable than in the quad cable: *in the coaxial cable crosstalk*

⁵⁾ See L. J. E. Kolk, Het balanceren van draaggolfkabels voor 48-kanalen systemen, PTT Bedrijf 3, 59-74, 1950 (in Dutch).

between pairs does not increase, but decreases with frequency.

This advantage can be accounted for by a consideration of skin effect. At sufficiently high frequencies, the signal currents flow nearly entirely on the outer surface of the inner conductor and on the inner surface of the outer conductor. The outer surface of the outer conductor carries practically no current, and between two points on this surface only very small alternating voltages will occur. When considering the influence of these voltages on an adjacent tube, the skin effect again is of importance: the density of the currents induced in that tube is greatest at the outer surface of the outer conductor, it sharply decreases towards the interior, which means that the interference is hardly capable of penetrating to the transmission circuit proper. This "shielding action" of the outer conductor of each tube is most effective at high frequencies. Thus, at sufficiently high frequencies, crosstalk will become entirely negligible.

The crosstalk shielding effect of the outer conductor not only permits the use of very high signal frequencies, but it also makes it possible to place the coaxial tubes, used for opposite directions of transmission, within a common lead sheath. Owing to the large difference in level between incoming and outgoing pairs at repeater stations, so exacting requirements as to crosstalk must be met with quad cables so that in this case separate "go" and "return" quad cables must, in general, be used.

From the above it is apparent that crosstalk on coaxial cables is only serious at the lowest frequencies transmitted. The lower frequency limit, therefore, is chosen fairly high. If it were desired to extend downwards the lower limit of the frequency band transmitted, this could be done by increasing

the thickness of the outer conductor, but only at the expense of increased cable cost. Some quantitative idea of this relationship may be obtained by a consideration of the depth of penetration ϑ of the signal and crosstalk currents into the outer conductor, i.e. the depth within the conductor at which the current density has diminished by a factor $1/e$. For a conductor having a resistivity ρ (in ohm metres) and a relative permeability μ_r , we may write:

$$\vartheta = \frac{10^5}{2\pi} \sqrt{\frac{\rho}{\mu_r f}} \text{ mm} = \frac{C}{\sqrt{f}} \text{ mm.} \quad (2)$$

In the above expression, the frequency must be expressed in kc/s. For copper it is found that $C \approx 2.1$; for aluminium $C \approx 2.75$. In order to reduce the depth of penetration to a few tenths of a millimeter, i.e. to an amount small compared to the minimum thickness of tape which can be applied without excessive mechanical difficulties, the frequency must exceed 100 kc/s. In coaxial systems on multitube cables, the lowest frequency transmitted is never less than 60 kc/s, often more ⁶⁾.

The upper limit of frequency in coaxial systems so far installed is approximately 2.8 Mc/s (660 channels).

Standardisation by the C.C.I.F. will in due course extend this band up to 4 Mc/s (see fig. 6), and work is now in progress on a transmission band up to 8 Mc/s. Naturally, at these increased frequencies the attenuation per km of cable is higher, and it will be necessary to use shorter repeater sections than have hitherto been customary. Nevertheless, as indicated above, this is not necessarily an economic dis-

⁶⁾ From equation (2) it will be seen that a high permeability results in a smaller penetration depth. For iron, $C \approx 0.3$. To save copper, the outer conductor therefore is sometimes covered with steel tape.

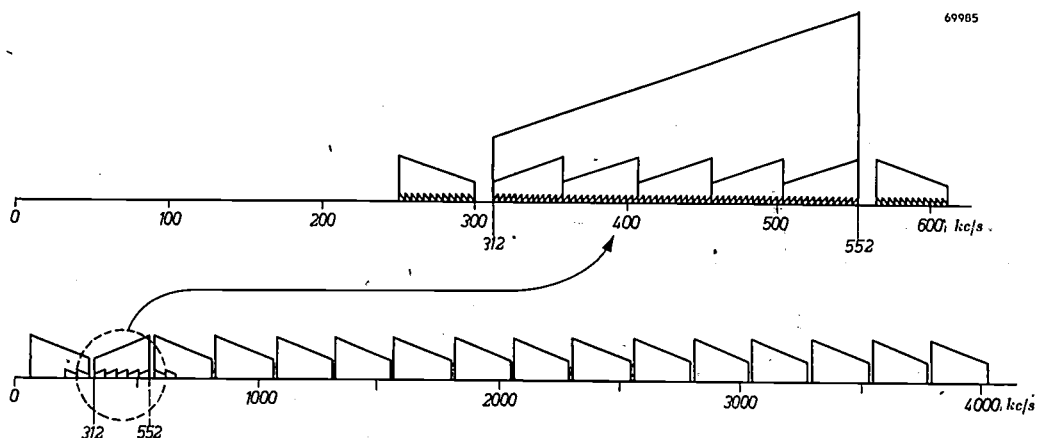


Fig. 6. Frequency allocation of a 960-channel coaxial cable system with 16 supergroups. The part within the dotted circle (supergroup no. 2) is enlarged to show more clearly the 5 groups per supergroup and the 12 channels (of nominal bandwidth 4 kc/s) in each group.

advantage, since the attenuation only increases as the square root of the frequency. For present types of coaxial cable the attenuation is given by the following formula:

$$a = \frac{0,07}{D} \sqrt{f} \text{ dB/km, (3)}$$

where f is again expressed in kc/s and D is the inside diameter of the outer conductor in cm. Fig. 7

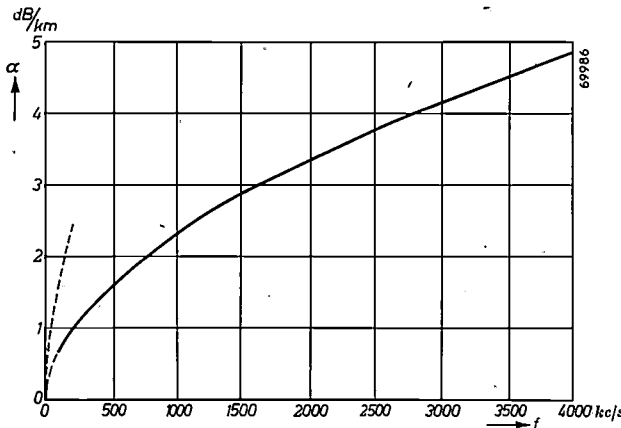


Fig. 7. Attenuation-frequency characteristic of a coaxial tube; D , the inner diameter of the outer conductor, is 0,94 cm. The broken line shows the corresponding attenuation of a 1.3 mm pair in a quad cable, reproduced from fig. 3.

shows the attenuation-frequency characteristic of a cable in which D has the standardised value 0,94 cm, mentioned above. The attenuation is 4.8 dB/km at 4 Mc/s. Assuming a repeater gain of 50 dB (which

allows a suitable margin), the necessary repeater spacing is thus about 9.5 km.

There is no economic objection to a considerably higher maximum frequency and a still smaller repeater spacing, but further progress in this direction is inhibited by technical difficulties in the repeaters.

Characteristics of repeaters for coaxial systems

Fig. 8 shows a comparison between two repeated cable systems of equal length, of the 12-quad (i.e. 24 pairs) and coaxial types respectively. "Go" and "return" paths are shown in each case. In the case of the quad cable it is assumed that each pair transmits 60 channels, i.e. 1440 in all, and that the repeater spacing is about 23 km. The coaxial tube accommodates 960 channels, the repeater spacing being taken as 9.5 km.

It is interesting to observe the saving in line amplifiers and physical circuit material per channel shown in the second diagram. Taking into account "go" and "return" transmission paths, the quad cable requires $2/(60 \times 23) \approx 1/700$ amplifier per channel km, whereas in the second case the corresponding figure is $2/(960 \times 9.5) \approx 1/(4500)$ amplifiers per channel km.

It is not our aim, however, once again to insist on the economic advantage of one system over the other. (To do this objectively we should have to consider the prices of cable and repeaters as well as those of the terminating equipment). We wish

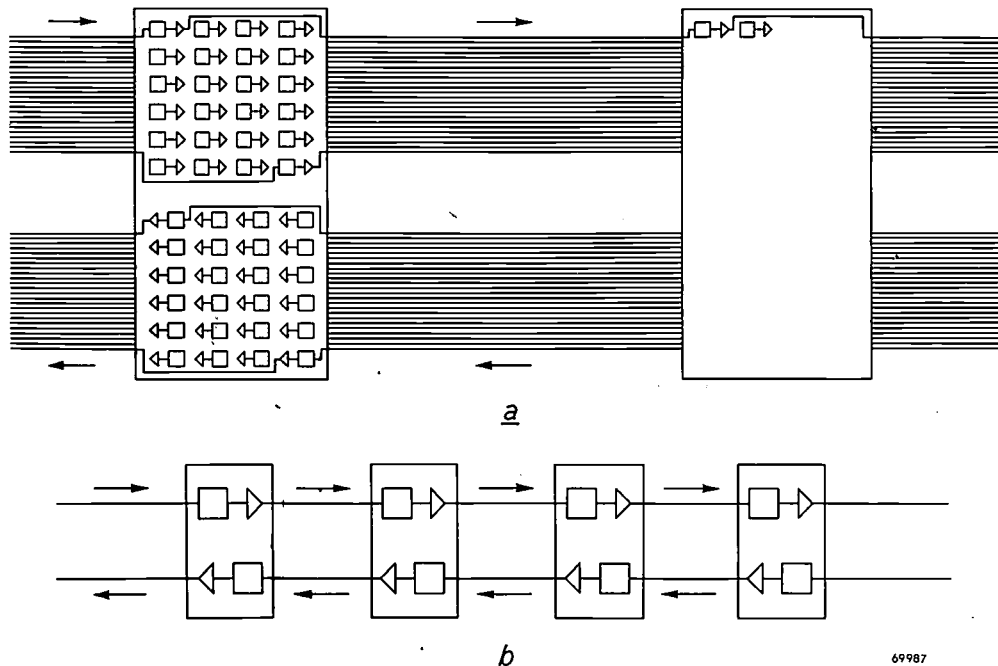


Fig. 8. a) Block schematic of go and return 12-quad cable with repeaters. Total number of channels is $60 \times 24 = 1440$.
b) Block schematic of a pair of coaxial tubes transmitting 960 channels.

rather to draw attention to some other features of this diagram:

- a) In the coaxial system each repeater handles 960 channels, whereas in the quad cable case, the corresponding figure is 60 channels.
- b) In the coaxial system each bi-directional physical circuit is equipped with 6 repeaters in a distance of 23 km, whereas in the quad cable system there are only 2.
- c) Finally, in the coaxial cable there are many small stations, and in the quad case, fewer large ones.

In consequence of these differences the line apparatus for coaxial cables in several respects must satisfy more stringent requirements than that used for quad cable. First, the reliability should be considered. The breakdown of one coaxial repeater would disable 960 channels simultaneously, which in telephony work would be regarded as little short of a catastrophe. Each repeater must therefore be provided with the necessary spare equipment, and it might even be advisable to provide a pair of coaxial tubes, fully equipped, as a standby.

Secondly, the stability is of importance. In a long-distance connection each channel passes through numerous repeaters in tandem. To maintain a constant level of all channels at the receiving end, variations of the individual gain of each repeater are strictly limited. Since there are more repeaters in tandem in a coaxial than in a quadded cable, for circuits of equal length, the limits for a coaxial system must be more stringent. This necessity is enhanced by the fact that the coaxial system is generally used for longer distances than quad cables.

Thirdly, non-linearity in the repeaters, especially the valves, gives rise to intermodulation, i.e. a form of (un-intelligible) crosstalk between channels. Since many channels contribute to intermodulation into any given channel, this effect is apparent as a background noise which is dependent upon the total number of channels per pair and upon the number of amplifiers in tandem. In both these respects coaxial cable again is at a disadvantage compared with quad cable.

The linearity of the amplifiers and their constancy of gain can theoretically both be improved by the use of negative feedback. In our case this requires a sacrifice of gain of 30-40 dB in each repeater. This, however, is particularly inconvenient because the stage gain of a broadband amplifier capable of handling 960 channels is already relatively small. This technical difficulty even sets a fundamental limit to the number of channels which can be transmitted over a single tube. In fact, the para-

sitic capacitances of amplifier valves, input and output transformers, etc., which are responsible for the decrease of stage gain at increasing bandwidth, also give rise to a phase displacement which increases with frequency; for this reason, there is a limit to the number of stages across which negative feedback may be applied; the limit in this case is three stages. It might be suggested that an amplifier could be built, for example, with two units, each consisting of three stages, with negative feedback applied to each unit individually. The sacrifice of 30-40 dB gain, however, would then apply to each such three-stage unit. Obviously, the placing of units in tandem will be of no use and no gain at all can be obtained if the available overall amplification per three-stage unit is already less than 30-40 dB. Beyond this limit any increase in the bandwidth transmitted over coaxial systems is thus conditional upon improvements in the characteristic of the components, particularly in respect of the capacitance and slope of the amplifier valves⁷⁾.

These, then, are the principal factors which for the moment set a limit to the increase of the frequency bands employed on coaxial cables.

The dispersal of the line apparatus over a large number of small stations, as demonstrated in fig. 8, is, unlike the subjects previously discussed, not primarily a circumstance which limits bandwidth; but an increase of bandwidth would certainly add to the practical difficulties of feeding power to a large number of repeater stations, possibly containing an increased number of valves. In quad cable systems it is customary to feed the line equipment from a local mains supply, a standby power supply being provided by a battery or engine-generator set. It is, however, not economical to furnish each of a large number of closely spaced coaxial repeater stations with its own local mains supply, nor with an emergency power source. The best solution of this problem so far devised is to feed each repeater station from a power supply transmitted over the cable itself. The necessary A.C. power is fed into the cable at stations spaced at intervals of up to 300 km along the cable route; these stations are provided with alternative sources of power supply, as a precaution against mains failure. By centralising the standby plant in this way, the problem of sup-

⁷⁾ For practical feedback and amplification values for a given bandwidth, see H. W. Bose, *Network analysis and feedback amplifier design*, Van Nostrand, New York, 1945. See also 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. In this connection it may be said that the use of Ferroxcube for the cores of input and output transformers in the amplifiers already represents an important step forward.

plying power to a large number of small stations has been solved in a very simple way and at a low cost per repeater. This system is all the more attractive since the solid copper core of the coaxial tube, which, with the exception of the extreme outer layer, does not serve the transmission of the signals, is then effectively utilized. Another advantage of the method is the easy solution it offers for feeding repeater stations in sparsely inhabited country where there may be no electric mains for hundreds of miles.

Another important problem in the line equipment, especially of coaxial systems, is the variation of α , the attenuation of the cable, as a function of temperature T . Evidently, the factor preceding \sqrt{f} in equation (3) and in which among other things the resistivity of the copper is involved, will vary with temperature. Now, da/dT will be obtained by multiplying the temperature coefficient of that factor — whatever function that may be — by \sqrt{f} . This means that fluctuations in the cable temperature will affect the level of the high-frequency channels considerably more than the low-frequency ones. The variations in level in coaxial systems have been reduced to reasonable proportions by providing an efficient form of automatic level regulation controlled by the line pilot signals. The extent of such variations for the reason explained being considerably less on quad cable systems, less elaborate regulation in that case is satisfactory.

The problem of interconnection

To complete our consideration of the relative merits of carrier systems on coaxial and quad cables, we shall now examine their respective adaptability to meet the demands of practical handling of carrier systems, in other words their "flexibility". Carrier telephone systems usually form part of a national or international telecommunication system. At various intermediate towns or junction points along the route a sufficient number of channels must be terminated to accommodate local traffic requirements. This number will vary over a wide range, and may not be constant. Temporary changes in the distribution of channels may be necessary on account of important events; or perhaps there may be steady growth of traffic due to an increase in the number of subscribers or the introduction of automatic operating procedure on trunk circuits. The structure of a trunk system is thus always changing and demands a certain degree of adaptability of the carrier system.

Systems incorporating 12-60 channels per pair such as those operated over quad cables, fulfil this condition admirably; by terminating a suitable number of pairs at each intermediate town or junction point, the distribution of channels can be readily adjusted to meet traffic requirements. On the other hand, the

same technique cannot be applied so easily to the coaxial system. Terminating one pair of coaxial tubes (go and return) at an intermediate town would mean providing for this town a hypergroup containing many hundreds of channels (potentially 960), i.e. much too large a unit for convenient use. This is the price which has to be paid to secure the economic advantages of transmitting a very large number of channels on a single physical circuit.

Some months ago this was reviewed in detail in this journal ⁴⁾ and reference was made to the possibility of interconnecting two coaxial systems on a "through group" or "through-super-group" basis, so that one hypergroup is no longer to be regarded as an indivisible unit, but some channels can be terminated at an intermediate point and the others can pass through in transit. The through super-group procedure requires that the incoming hypergroup be demodulated to its component basic supergroups, whose frequency range is 312-552 kc/s. The supergroups which terminate are demodulated completely to audio frequency, whereas those in transit pass via "through-super-group filters" to the super-group modulators, to take their place in the outgoing hypergroup.

Although in principle this arrangement imparts to the coaxial system the flexibility of 12- or 48-channel quad cable systems, the additional equipment at the transit point inevitably must detract from the economic advantage of the coaxial system. An alternative and simpler method of providing transit facilities exists; this is known as echelon working and it consists in allocating to the traffic between each pair of towns in a coaxial cable network the exclusive use of one or more of the supergroups provided by a pair of tubes. The intermediate carrier terminals then only need be connected in parallel with the main transmission paths by any convenient method, without any of the special filters mentioned above. Admittedly the available bandwidth in general is not fully utilised when this scheme is adopted, and consequently part of the essential feature of the coaxial cable system is sacrificed.

A completely different solution of the problem that is being studied involves a rearrangement of the supergroups with a wider mutual separation. This proposal will usually result in a less efficient utilisation of the frequency bandwidth, but will permit of a new technique, the use of "band-stop filters" to suppress one or more of a range of supergroups for purposes of extraction at intermediate terminals. This would greatly simplify the distribution problem ⁴⁾.

The above discussion indicates that coaxial and quad cables each have their own characteristics which adapt them for specific applications, neither type having a decisive advantage over the other in all fields. Future developments may eventually lead to one system gaining ground at the expense of the other. On the basis of present technique, the coaxial system undoubtedly is to be preferred for projects in which a large number of channels must be provided over long distance, and in which the characteristic economy in the cost of cable and repeaters is consequently obtained. In sparsely populated areas where power supplies are mostly non-existent, the coaxial system has the important advantage of power feed over the cable; in some areas, e.g. in the U.S.A., this advantage has been decisive. For comparatively short distances in thickly populated country, particularly where the cable network forms a closely triangulated pattern, quad cable systems with relatively few channels (e.g. 48) have generally been preferred on account of their greater flexibility. There are many marginal cases between these two extremes; the ultimate choice will here depend on the progress of future development, for example the increase in frequency bandwidth of quad cable mentioned above, and the proposal to feed power over quad cable to repeater stations. These expected improvements will then have to be compared with the corresponding evolution of coaxial technique.

Finally, it should be mentioned that the feasibility of relaying television programmes is being advanced as an important argument in favour of coaxial cable. For this purpose a bandwidth of at least 4 Mc/s is required. A combination of telephone and television transmission over the same medium would undoubtedly be attractive.

Appendix: Transmission characteristics of coaxial cable

The transmission characteristics of any line, for a signal of frequency $f = \omega/2\pi$, are determined by its resistance R , inductance L , leakage G , and capacitance C , which properties are continuously distributed along the line and which are expressed per unit length. The values of R , L , G and C , the "primary constants" are usually functions of frequency, in spite of their name, but only to a slight extent in the case of L and C .

The theory of transmission becomes fairly simple when the line is uniform and when $\omega L \gg R$ and $\omega C \gg G$, which is always the case in practice at high frequency, i.e. well beyond the audio range. According to theory, in that case the ratio of signal voltage E to signal current I , both assumed sinusoidal and transmitted into an infinitely long line, is the same in all points and equal to $\sqrt{L/C}$; this ratio is independent of frequency and is called the characteristic impedance. The losses in transmission are caused by the resistance R and the leakage G . If the current and voltage transmitted into an infinite line are I_1 and E_1 , at unit length from the send-

ing end the smaller values I_2 and E_2 are found. The attenuation of the cable is defined by the following equations:

$$a = \log_e |E_1/E_2| = \log_e |I_1/I_2|.$$

From theory, the following simple formula for a is obtained:

$$a = \frac{R}{2} \sqrt{\frac{C}{L}} + \frac{G}{2} \sqrt{\frac{L}{C}} \dots \dots \dots (4)$$

It contains two terms, corresponding with the two distinct causes of loss: the first term represents resistive attenuation due to R , and the second term, always considerably smaller, the leakage attenuation due to G .

So far the theory applies to all types of line. In order to apply formula (4) to our case, we need to relate the primary constants to the dimensions and material constants of the coaxial cable.

Consider first the resistance R . The high-frequency current chiefly flows in a thin cylindrical layer at the outer periphery of the inner conductor (diameter d) and at the inner periphery of the outer conductor (diameter D). Now R can be taken as equal to the D.C. resistance of two tubular conductors with the diameters d and D respectively and a wall thickness equal to the penetration depth δ . The cross-sectional areas of these conductors are $\delta \cdot \pi d$ and $\delta \cdot \pi D$, and according to eq. (2) we have $\sqrt{\rho_1/f}$ and $\sqrt{\rho_2/f}$ respectively, ρ_1 and ρ_2 being the resistivities of the inner and outer conductor material respectively. The resistance R then is proportional to

$$\sqrt{f} \left(\frac{\sqrt{\rho_1}}{D} + \frac{\sqrt{\rho_2}}{d} \right).$$

The capacitance C as well as the inductance L of two coaxial cylinders depends on the ratio of the diameters only, not on their absolute values. It is found that $\sqrt{L/C}$, the characteristic impedance, is proportional to $\log D/d$. The resistance part of the cable attenuation given in equation (4) then becomes proportional to

$$\frac{\sqrt{f}}{D} \frac{\sqrt{\frac{\rho_2}{\rho_1}} + \frac{D}{d}}{\log \frac{D}{d}} \dots \dots \dots (5)$$

The second fraction in this expression depends on the ratio D/d and for a certain value of this ratio it passes through a (fairly flat) minimum. This optimum ratio is $D/d = 3.6$ for $\rho_2/\rho_1 = 1$, the usual case (inner and outer conductors made of same material). Thus, if we regard D as fixed, the cable attenuation will attain a minimum if d is made equal to $D/3.6$, in accordance with the condition (1). When the outer conductor is of aluminium and the inner of copper, $\rho_2/\rho_1 = 1.64$ and the condition of minimum attenuation occurs when $D/d = 3.8$.

Substituting for the second fraction in (5) the minimum value $4.6/\log_e 3.6$ and, taking into account the proportionality factors, not embodied for simplicity's sake in the above formulae, we finally obtain equation (3) for the attenuation. (In this case it is assumed that both conductors are of copper, for which $\rho = 0.01745 \times 10^{-6}$ ohm metres, and that the dielectric is air ($\epsilon_r = 1$.) By choosing D/d , likewise the inductance L and, when the relative dielectric constant ϵ_r of the insulation in the cable is known, the capacitance C and the characteristic impedance $\sqrt{L/C}$ of the coaxial tube are fixed. The characteristic impedance for a cable of optimum design with air dielectric is 77 ohms. With a typical construction using spacer beads or the equivalent, this figure is reduced to 75 ohms and with solid polythene insulation to 52 ohms.

The leakage G , which is caused by dielectric losses in the insulating medium, may be expressed in the form of a loss angle δ for the dielectric, by the equation:

$$G = \omega C \tan \delta. \quad \dots \dots \dots (6)$$

With most insulating materials used for coaxial cables, $\tan \delta$ does not depend on frequency within the frequency range of interest.

Since C , except for being proportional to ϵ_r , only depends on the ratio D/d , which has already been chosen for optimum resistance attenuation, it follows that, with this optimum design, G will be proportional to $f \cdot \epsilon_r \cdot \tan \delta$. For polythene ϵ_r is not large (about 2-3) and dielectric losses are extremely small: a value $\tan \delta \approx 3 \times 10^{-4}$ is usually assumed. This leads to a leakage attenuation component which may be disregarded in ordinary cable construction, notwithstanding the fact that this component increases with frequency at a greater rate than the resistance attenuation.

Summary. Development of carrier telephone systems has been directed towards increasing the number of channels per

physical circuit up to the economic limit in order to reduce the line cost per channel. In "quad" cables, 48 channels per pair is now normal practice (maximum frequency transmitted 204 kc/s). It is hoped to extend this figure in the future to 60, 72 or even 120 channels per pair. Further progress in quad cable development is limited by crosstalk which becomes more serious at the higher frequencies. On this score, things are fundamentally better with coaxial cable; owing to the screening effect of the outer conductor, crosstalk in this case decreases with increasing frequency. Crosstalk at frequencies below 60 kc/s is serious on coaxial cables and consequently this is normally the lowest frequency transmitted, but there is in this respect no upper frequency limit. Coaxial systems at present in operation in Great Britain and the U.S.A. transmit as many as 600-660 channels per pair, and standardisation by the C.C.I.F. provides for a future increase to 960 (maximum frequency transmitted 4 Mc/s). An economic limit is imposed by level considerations which lead to a reduction in the repeater spacing not commensurate with the number of extra channels obtained. This economic limit has not yet been reached, the number of channels being actually restricted due to the aggravation with increasing frequency of requirements in respect of reliability, stability and linearity of the repeaters. Finally, the problem of interconnection is discussed, this being inherent in all carrier systems having few physical circuits each carrying a large number of channels.

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.

1995: J. J. Went: The value of the spontaneous magnetization of binary nickel alloys as a function of temperature (*Physica* 17, 596-602, 1951, No. 6).

The J_s versus T curve for pure nickel is more concave towards the T -axis than that for any binary nickel alloy, with only one exception, viz. completely ordered alloys, such as slowly-cooled Ni_3Fe . For pure nickel the form of the J_s versus T curve can be explained by the occurrence of an order-disorder phenomenon of the magnetic moments, where the latter can be placed only parallel or anti-parallel. It is suggested that all the other J_s versus T curves must be explained as being a result of this order-disorder phenomenon and of the statistical fluctuations of the concentration of dissolved atoms. In the case of Ni_3Fe two different order-disorder phenomena (a crystallographic and a magnetic one) act simultaneously.

1996: J. Smit: Magnetoresistance of ferromagnetic metals and alloys at low temperatures (*Physica* 17, 612-627, 1951, No. 6).

The magnetoresistance of pure Ni and Fe, of Ni-Fe-, Ni-Co-, and Ni-Cu-alloys and of Heusler's alloy has been measured at room temperature and at temperatures of liquid nitrogen and liquid hydrogen. The behaviour of the pure metals is

essentially different from that of the alloys. A maximum in the magnetoresistance is observed at low temperatures for alloys having about one Bohr magneton per atom. The positive difference between the longitudinal and the transversal resistance can be explained by means of the spin-orbit interaction.

At low temperatures the pure metals show an increase in resistance with increasing field just as the non-ferromagnetic metals. From this the value of the internal field, acting on the conduction electrons, could be determined, and was found to be approximately equal to the flux density B .

1997: J. A. Haringx: De instabiliteit van inwendig op druk belaste dunwandige cylinders (*De Ingenieur* 63, O 39-O 41, 1951, No. 29). (The instability of thin-walled cylinders subjected to internal pressure; in Dutch.)

It is shown that under certain conditions a thin-walled cylinder may buckle when subjected to internal pressure. The critical value of this pressure can easily be calculated on account of Euler's well-known formula. Most of the formulae for pressure-loaded cylinders given in current textbooks, however, fail to predict this behaviour and should therefore be applied with caution for great lengths and/or for high pressures.

1998: J. A. Haringx: De instabiliteit van inwendig op druk belaste ronde balgen (De Ingenieur 63, O 42-O 44, 1951, No. 29). (The instability of cylindrical bellows subjected to internal pressures; in Dutch.)

Like thin-walled cylinders, dealt with in a previous paper (see No. 1997), also bellows may become unstable when loaded by internal pressure. The critical value of this pressure, which is governed, through Euler's well-known formula, by the rigidity of the bellows with respect to bending, is computed only for rectangularly shaped corrugations. The result found has been checked experimentally.

1999: C. G. Koops: On the dispersion of resistivity and dielectric constant of some semiconductors at audiofrequencies (Phys. Rev. 83, 121-124, 1951, No. 1).

Semi-conducting $\text{Ni}_{0.4}\text{Zn}_{0.6}\text{Fe}_2\text{O}_4$, prepared in different ways, has been investigated. It appeared that the A.C. resistivity and the apparent dielectric constant of the material shows a dispersion, which can be explained satisfactorily with the help of a simple model of the solid: there should be well-conducting grains, separated by layers of lower conductivity. Dispersion formulas are given. There is good agreement between experiment and theory.

2000: E. J. W. Verwey: Oxidic semi-conductors (from: Semi-conducting materials, Butterworth's Scientific Publications Ltd., London 1951, pp. 151-161).

Non-conducting oxidic materials may be rendered conductive either by preparing solid solutions with oxides of high conductivity (e.g. a poorly conducting spinel with Fe_3O_4) or by introducing ions of deviating valency into the lattice (method of controlled valency). A comparison is made between semi-conductors obtained in this way and elemental semi-conductors such as silicon and germanium. The conductive properties of polycrystalline semi-conductors as a function of frequency are discussed, especially the influence of non-conducting thin layers at the grain boundaries (see these abstracts, Nos. 1738, 1739, 1845, 1914, and Philips techn. Rev. 9, 36-54, 239-248, 1947; 13, 90-95, 1951, No. 2).

2001: J. Volger: Some properties of mixed lanthanum and strontium manganites (from: Semiconducting materials, Butterworth's Scientific Publications Ltd., London 1951, pp. 162-171).

Discussion of electrical properties of manganites XMnO_3 . This material is an example of a

polycrystalline semi-conductor prepared according to the principle of controlled valency (replacing the trivalent ion X, e.g. La, by a bivalent ion, such as Sr). There exists a close analogy between the resistivity and the ferromagnetic properties. Discussion of anomalies at the Curie temperature, of the influence of frequency, electric field and external magnetic field on the resistivity, of the thermo-electric properties and the Hall effect.

2002: H. Bremmer: On the diffraction theory of Gaussian optics (Comm. pure and appl. Math. 4, 61-74, 1951, No. 1).

The diffraction theory of optical imaging is developed for objects with arbitrary structure. The theory is based on rigorous solutions of the wave equation instead of the conventional approximation of Kirchhoff's formula. The similarity of the wave functions in the object plane and in the corresponding paraxial image plane (Gaussian systems with unlimited aperture) proves to be connected with Neumann's integral theorem for Bessel functions (instead of the Fourier identity as in Kirchhoff's approximations). Another solution accounts for the effects of optical aberrations and of limited apertures.

2003: H. Bremmer: The W. K. B. approximation as the first term of a geometric-optical series (Comm. pure and appl. Math. 4, 104-115, 1951, No. 1).

The W.K.B. approximation of the solution of $y'' + k^2(x)y = 0$ is derived from a discontinuous model of an inhomogeneous medium. Higher approximations are found by considering multiple reflections. The solutions of different order form a series, the convergence of which is discussed. The well-known insufficiency of the W.K.B. approximation in the neighbourhood of zero's of $k(x)$ can be interpreted as a very slow convergence of the series mentioned.

2004: W. Ch. van Geel: On rectifiers (Physica 17, 761-776, 1951, No. 8).

Experiments are described in which rectifiers were obtained from combinations of metals, semi-conductors and intermediate layers. The following combinations have been investigated: (a) the contact between an excess semiconductor and a deficit semiconductor; (b) the combination aluminium/aluminium oxide/semiconductor; and (c) the combination metal/resin layer/semiconductor. In all these cases rectification occurs. The suggestion

is put forward that, in all three cases, the contact between two layers with charge carriers of opposite sign is the cause of rectification.

R 173: J. C. Francken and R. Dorrestein: Paraxial image formation in the "magnetic" image iconoscope (Philips Res. Rep. 6, 323-346, 1951, No. 5).

In this paper a method is described for computing paraxial rays in a cathode lens placed in a magnetic field. In order to judge the approximations made, the well-known derivation of the ray equation is given. The solutions of this equation are discussed and a physical interpretation is given. A special case, approximating the electron-optical system in the image iconoscope, is computed numerically. The electron trajectories thus found lead to a discussion of the imaging mechanism in these electron-optical systems. In some respects the mechanism appears to differ considerably from that in ordinary magnetic lenses.

R 174: G. J. Fortuin: Visual power and visibility, II, (Philips Res. Rep. 6, 347-371, 1951, No. 5).

See **R 170**. This part deals with the distribution of the visual power and of the visual types in age groups. Further the physiological meaning of the constant factors in the quantitative definition of visual power is discussed. One of these factors represents the lowest field brightness that allows perception of a dark object, another represents the brightness at which rod vision changes into cone vision and vice versa.

Visibility is the subject of the last chapter. Three possible definitions were tested by some standards, but only the ratio between the actual size of the object and the size of the smallest object perceptible under equal conditions (the so-called size-reduction factor) complied with our demands.

R 175: J. L. H. Jonker: The angular distribution of the secondary electrons of nickel (Philips Res. Rep. 6, 372-387, 1951, No. 5).

The common equipment for measuring secondary emission (disc-shaped or spherical electrode within a sphere) is not suitable to obtain data about the angular distribution of the secondary electrons. To this aim an electrode system with two concentric spheres was constructed in order to obtain a really radial retarding electrostatic field, with which the behaviour of the secondary electrons with different velocities could be studied. The distribution of the secondary electrons (slow genuine secondary electrons, secondary electrons with moderate velocities, and rapid reflected electrons) was measured as a function of the angle of incidence and of the bombardment voltage of the primary electrons. The construction of the measuring tube, the method of measuring and the results obtained are discussed.

R 176: N. Warmoltz: The time lag of an ignitron (Philips Res. Rep. 6, 388-400, 1951, No. 5).

To ignite periodically a discharge in a rectifier with a mercury-pool cathode, a current is sent through a semiconductive rod partly immersed in the mercury. The time lag in starting the discharge is measured on an oscillograph when a charged capacitor is switched by a thyatron onto the igniter. This is done for igniters of widely varying resistance in a liquid and a solid mercury cathode and in a liquid and a solid tin cathode. Also the influence of the gas pressure in the tube is investigated. The results are compared with those of Slepian and Ludwig and those of Dow and Powers. Finally the field emission theory of Slepian and Ludwig and the thermal theory of Mierdel are discussed, whereby it turns out that the measurements of the time lags with liquid and solid cathodes fit best in the thermal theory.

ERRATUM

The colour reproduction of a painting by Giovanni Bellini on page 77 in the preceding issue of this Review is by courtesy of the Trustees of the National Gallery, London. The editors regret that owing to a mistake this statement was omitted in the subscript.

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 NOISE OF ELECTRONIC VALVES AT VERY HIGH FREQUENCIES

I. THE DIODE

by G. DIEMER and K. S. KNOL.

621.396.822:621.396.645.029.6

Noise of radio and television receivers, in so far as it is produced in the receiver itself, is caused by a number of rather complicated phenomena. These phenomena can nevertheless be investigated fairly well theoretically at wavelengths that are not too small. However, at wavelengths within the decimetric and particularly centimetric ranges the noise problem becomes more difficult. The transit times of the electrons passing through electronic valves are then no longer small compared with the oscillation of the alternating electric field, so the waveform of the individual current pulses plays a part. This article explains the complications this leads to in the analysis of the noise phenomenon of a diode. In another article to be published later, noise phenomena in triodes will be investigated at very short wavelengths. It will be shown then how noise is affected by the nature of the circuit in which the valve is employed.

The term "noise", originally derived from acoustics, is used in electrical engineering in a very general sense. Here it is used to denote random fluctuations, for instance of currents and voltages. Such fluctuations may be of thermal origin — thermionic noise — or they may be connected with the corpuscular structure of electricity. In electronic valves the number of electrons passing through a certain cross section of a valve per unit time is subject to variations, and this causes fluctuations in the current flowing through the output circuit.

The causes of the noise which occurs, for instance in radio and television receivers and amplifiers at high frequencies (metric and decimetric waves), have been dealt with in previous issues of this journal¹⁾. Since then, however, the frequency range that can be used for practical purposes has been extended to cover centimetric waves, and this has given rise to all sorts of specific problems, connected with noise. It is these problems that will be

treated in the present article. Following upon a general introduction, attention will be directed particularly to the noise of a diode. Another article is to follow, dealing with the noise of a triode and the manner in which noise is affected by the circuit in which the valve is working.

The causes of noise in radio and television receivers can be classified under three headings: those outside the receiver (aerial noise), those contained in the electronic valves of the receiver (electronic noise) and those in the other parts of the receiver (network or Johnson noise). Since our intention is to investigate the electronic noise in particular, the other two categories will be reviewed only briefly. Further information about them can be found in the literature referred to in footnote¹⁾.

Aerial noise and network noise

Aerial noise

The primary contribution to aerial noise is made by electrical apparatus in the vicinity of the receiver. The ignition mechanisms of motor vehicles, gas-discharge lamps, electric motors and suchlike

¹⁾ M. Ziegler, Philips techn. Rev. 2, 136-141, 1937; 2, 329-333, 1937, and 3, 189-196, 1938; C. J. Bakker, Philips techn. Rev. 6, 129-138, 1941; M. J. O. Strutt and A. van der Ziel, Philips techn. Rev. 6, 178-185, 1941.

are sources of noise known to many of us from experience. Electric discharges (sparking, for instance) always result in electromagnetic radiation over a wide frequency range, and it is this radiation that manifests itself as noise. It can often be counteracted fairly well at its source and is often called man-made interference, to distinguish it from other phenomena to be dealt with below.

Electric discharges also take place in the atmosphere, namely the discharges of electric storms (lightning). Electro-magnetic radiation with a wavelength greater than about 20 metres is reflected back to earth by the ionosphere, so that atmospheric noise in this wave range may become noticeable at very great distances from its source. The electric storms which occur so frequently in the tropics can thus cause atmospheric noise over the whole of the earth. On wavelengths less than 10 to 20 metres, however, the effect of atmospheric noise is less noticeable.

There is also electro-magnetic radiation of a cosmic origin (galactic noise). In recent years such radiation has proved of great importance in astronomy, but here we are concerned solely with the fact that it acts as a source of noise in receivers. Since the ionosphere transmits only radiation with a wavelength less than about 20 metres, cosmic noise is troublesome on the short waves only. It is strongest on the wavelengths between 1 and 10 metres, where cosmic noise is comparable to atmospheric noise and the noise of the receiver itself. In the construction of receivers for wavelengths in this wave range, therefore, it serves little purpose to pay very much attention to receiver noise. On the ultra short waves (below 1 metre) the intensity of cosmic radiation is so small that little trouble is experienced from it; in this region, therefore, it is important to restrict receiver noise as far as possible.

Network noise

From thermodynamic considerations it can be deduced that every resistance, R , occurring in a circuit acts as a source of noise. It is equivalent to a current generator with parallel conductance $1/R$, producing a fluctuating short-circuit current i . The mean square value of i in the frequency range between f and $f + \Delta f$ is given by:

$$\bar{i}^2 = \frac{4kT}{R} \Delta f, \dots \dots \dots (1a)$$

where T represents the absolute temperature of the resistor and k is Boltzmann's constant. This expression is not frequency dependent. The resistor may also be regarded as a noise-voltage generator

with internal resistance R , yielding a fluctuating e.m.f. u . The mean square value of u in the frequency range between f and $f + \Delta f$ is given by

$$\bar{u}^2 = 4kTR\Delta f. \dots \dots \dots (1b)$$

The resistor is, of course, coupled to the rest of the circuit, which must be regarded as a load on the noise generator. Assuming for a moment that the resistor and the rest of the circuit are at the same temperature, there is just as much noise power passing from the resistor to the circuit as in the opposite direction. This noise power is a maximum when the internal resistance of the circuit is equal to R (optimum matching). As a time average the noise power in a frequency range Δf is then equal to

$$\bar{p}_{\max} = kT\Delta f. \dots \dots \dots (2)$$

If the temperature of the resistor differs from that of the circuit, which is, let us say, at room temperature T_0 , then the thermodynamic equilibrium on which the above considerations are based is broken; there is no longer the same amount of power passing in both directions. With optimum matching, however, the noise power produced by the resistor is still equal to the value given by formula (2). The ratio of T to T_0 is called the noise factor F of the resistor. The factor F is a ratio of two powers and is often expressed in decibels, being called then the noise figure (N). It appears that the conception of a noise factor can easily be extended to the case where a non-thermal source of noise is present.

Different parts of the noise-producing resistor (or, more generally, the noise-producing network) may have different temperature. Then a certain average T^* of the temperatures occurring in the network, the equivalent noise temperature, can be defined such that

$$\bar{p}_{\max} = kT^*\Delta f. \dots \dots \dots (3)$$

F is then given by T^*/T_0 .

Deviations from thermodynamic equilibrium may also arise in another way, when the circuit contains sources of macroscopic voltages giving rise to macroscopic currents in the resistors²⁾. In metallic resistors, even when the current is large, there is only a very slight disturbance of the equilibrium, so the noise does not change (the drift velocity of the electrons is small compared with their thermic velocities). In semi-conductors, carbon resistors or

²⁾ By macroscopic quantities are meant those quantities which are not subject to random fluctuations, such as, for instance, ideal battery and signal voltages. These quantities will be indicated in capital letters, to distinguish them from noise quantities (which might also be called microscopic quantities) written in small letters.

contact resistors, however, a macroscopic current may cause a considerable increase in the noise (apart from any increase due to a rise in temperature). This noise appears to be frequency dependent, being greatest usually at low frequencies.

Electronic noise at frequencies up to 10^7 c/s

A thermodynamical approach has been used, so far, in describing the causes of noise in electronic valves. It is equally possible, however, to describe them in terms of the corpuscular structure of an electronic current. There is one type of valve, that having a double cathode³⁾, in which the electrons are in thermodynamic equilibrium with the surroundings; the same result is reached in this case, whichever method of reasoning is followed. In other types of electronic valves, however, great deviations from thermodynamic equilibrium commonly occur, so in this article the second method of reasoning will be followed, except where the laws of thermodynamic equilibrium are obviously applicable.

The problem of noise in electronic valves is already rather a complicated one, but for very high frequencies (10^9 - 10^{11} c/s) it becomes still more difficult, since the transit times of the electrons are comparable to the cycle of the alternating field. We shall therefore consider first electronic noise at medium frequencies, in the range of the decametric waves. This has already been discussed at length in the articles quoted in footnote¹⁾, so here a summary will be given, supplemented by results of more recent investigations in this field.

Shot effect

The number of electrons passing through a certain cross section of a valve per unit is subjected to fluctuations owing to statistical fluctuations in the number of electrons emitted by the cathode per unit time. It can be proved that the noise of a saturated diode — that is a diode in which the anode is at such a high potential that all electrons leaving the cathode are attracted to the anode — is caused entirely by shot effect. The fluctuations in anode current are expressed, as before, by the mean square value of the amplitude of the noise current in the frequency range between the frequencies f and $f + \Delta f$. They are given by:

$$i^2 = 2eI_a\Delta f, \dots \dots \dots (4)$$

where I_a stands for the anode direct current and e

is the absolute value of the electron charge. This formula for the noise due to shot effect applies for all valves, subject only to the condition that the electrons move towards the anode independently of each other, as they do, for instance, in the exponential range of a diode. If, there is space charge, formula (4) no longer holds.

Saturated diodes are often used as noise standards. For not too high frequencies the internal parallel conductance is zero ($dI_a/dV_a = 0$, the characteristic of saturation), so that the diode is an ideal current generator. For frequencies at which the finite transit time of the electrons begins to play a part, formula (4) is no longer correct, for the noise then decreases in intensity. A diode which serves as a noise standard as far as the decimetric-wave range is the Philips type 10 M diode.

Effect of space charge (noise suppression)

Epstein has shown that the potential in a diode at the average anode voltage varies in the manner represented in fig. 1. There is a minimum in the

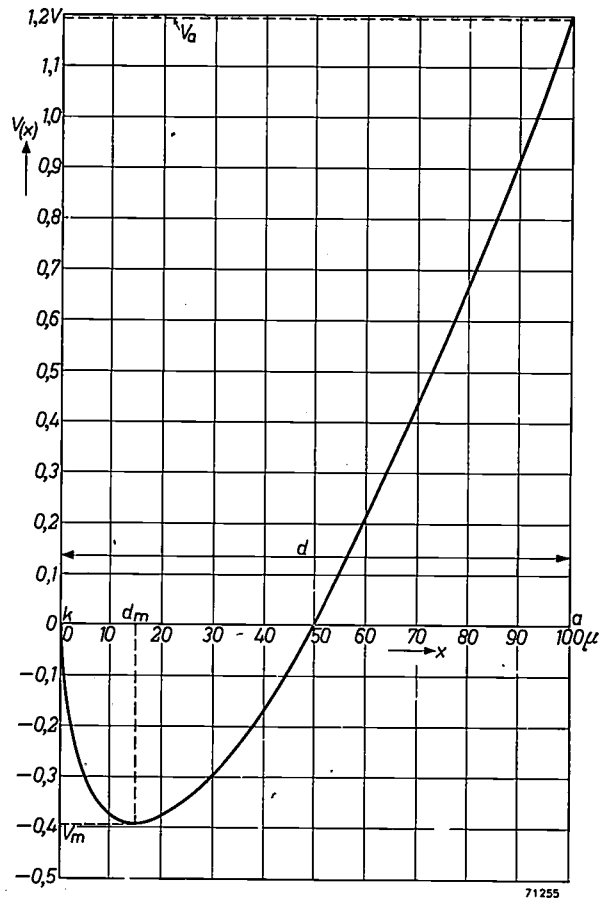


Fig. 1. Potential distribution in a diode with space charge, for a practical case. The potential $V(x)$ is plotted as a function of the distance x from the cathode k (imagined as being at a potential 0); d is the distance between the electrodes, V_a the anode voltage. At a distance d_m from the cathode is a minimum in the potential curve, with depth V_m . The potential minimum is the most difficult part of the valve for the electrons to pass through (owing to their negative charge).

³⁾ D. K. C. MacDonald, Proc Roy. Soc. A 195, 225-230, 1948; K. S. Knol and G. Diemer, Philips Res. Reports 5, 131-152, 1950.

potential curve near the cathode (at a distance d_m), this being due to the space charge formed by all the electrons which at a certain moment are present between the cathode and the anode. This potential plot follows from theoretical considerations⁴⁾. Experimentally it can be checked only indirectly, owing to various disturbing effects making the results of the experiment rather uncertain (variations in the work function across the electrode surfaces, secondary anode emission and suchlike). It is not, therefore, to be expected that the following considerations will show any exact numerical agreement with what is found in practice. The depth of the potential minimum depends upon several factors, amongst them the anode voltage and the cathode temperature; if the anode voltage is very high or the cathode temperature very low, there is no minimum.

The electrons do not all leave the cathode at the same velocity. The average number of electrons $\bar{N}(v_x)dv_x$ having a forward velocity component between v_x and $v_x + dv_x$ is given by the distribution function:

$$\bar{N}(v_x)dv_x = \frac{I_s m}{ekT} e^{\frac{mv_x^2}{2kT}} v_x dv_x, \dots (5)$$

where k and T have the known meaning, m stands for the mass of the electron and I_s is the saturation current of the cathode. Only the fastest electrons are able to pass the potential minimum and contribute towards the anode direct current. $N(v_x)$ fluctuates around the value given by formula (5). If the amplitude of the fluctuations in the current reaching the anode were to be relatively equal to that of the fluctuations in the current leaving the cathode, then formula (4) would apply also when there was space charge. The fluctuations in the number of electrons passing from the cathode to the anode are, however, relatively smaller than those occurring in the total number of electrons leaving the cathode, so the noise is proportionately weaker. The reason for this is that a temporary increase of the number of electrons emitted makes the potential minimum so much deeper that some of the electrons — those whose velocities lie within certain narrow limits (see below) — are unable to reach the anode. Thus the noise is partially suppressed by the space charge. This is expressed by the formula:

$$\bar{i}^2 = \Gamma^2 2eI_a A f, \dots (6)$$

where Γ represents the noise-suppression factor: $\Gamma < 1$.

⁴⁾ P. H. J. A. Kleijnen, Philips Res. Rep. 1, 81-96, 1946; A. van der Ziel, Philips Res. Rep. 1, 97-118, 1946.

For low frequencies it is possible to calculate the value of this factor Γ^5). In fig. 2 both calculated and experimental values of Γ^2 are given as functions of the anode voltage. The noise has been measured on a triode connected as diode and to be regarded as a diode with a very low secondary emission.

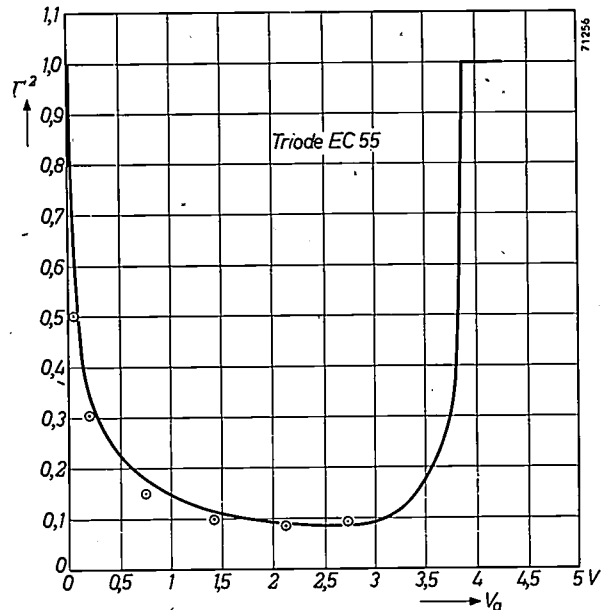


Fig. 2. The square of the noise-suppression factor Γ at the frequency $6 \cdot 10^6$ c/s as a function of the anode voltage V_a . This diagram applies for the triode EC 55 connected as a diode. The curve represents the theoretical trend, while the small circles are the values found experimentally.

Reflection noise

Electrons reaching the anode may be reflected back by the surface of the anode and they may also release secondary electrons from the anode. The reflected electrons are the fastest, possibly penetrating into the region of the potential minimum and thus increasing its depth; the secondary electrons are much slower and as a rule have little effect. This effect of the reflected electrons reduces the anode direct current but increases the noise intensity. The number of returning electrons fluctuates in time. Denoting the average reflected-electron current passing from the anode to the potential minimum by I_r and the electron current reaching the anode by I_a , then:

$$I_r = rI_a,$$

where r is the reflection coefficient. According to (4) the fluctuations in I_r will be given by:

$$i_r^2 = 2eI_r A f = 2erI_a A f.$$

For the sake of simplicity it will be supposed that all the reflected electrons reach the cathode. An upper limit may then be found for their effect upon the noise: I_r must be subtracted from I_a and i_r added to i_a . Owing to the deepening of the potential minimum, fluctuations arise in I_a which are practically equal to i_r . At low frequencies these fluctuations are in

⁵⁾ W. Schottky and E. Spenke, Wiss. Veröff. Siemens Werke 16^{II}, 1-41, 1937; D. O. North, B. J. Thompson and W. A. Harris, R.C.A. Rev. 4, 269-289 and 441-472, 1940.

the same phase as those in I_r , so that the total fluctuations in I_a due to reflected electrons is given by:

$$\overline{(2i_r)^2} = 8eI_aAf.$$

The shot effect in the anode current gives rise to the fluctuations expressed by formula (6), so the total noise is given by:

$$i^2 = \overline{i_a^2} + \overline{(2i_r)^2} = 2eI_a(\Gamma^2 + 4r) \dots (7)$$

If, for instance, $\Gamma^2 = 0.05$, which is a fairly normal value, an anode reflection coefficient of 1% is already sufficient to double the intensity of the noise, as is illustrated in fig. 3.

Secondary-emission noise

A surface specially prepared for the purpose can act as electron multiplier through secondary emission. The number of secondary electrons released by a primary electron varies statistically: for a given surface there is an emission function $\beta(m)$ indicating the chance of a primary electron releasing m secondary electrons. The mean value of m is called the secondary-emission coefficient δ of the surface, δ being given by:

$$\delta = \overline{m} = \Sigma\beta(m)m.$$

For a good secondary-emitting surface the coefficient δ will be about 5.

Owing to the fluctuations in m , the secondary-emission current I_{sec} generated by a primary current I_{pr} (for the present imagined as being free of fluctuations) will cause noise. According to the rule of probability the spread in the number of secondary electrons per primary electron is given by:

$$\sigma_{sec}^2 = m^2 - (\overline{m})^2 = \kappa\delta - \delta^2 (> 0),$$

where $\kappa\delta$ is substituted for $\overline{m^2} = [\Sigma\beta(m)m^2]$ in order to arrive at an expression that can be conveniently handled.

The mean secondary direct current I_{sec} is equal to δI_{pr} , and from the foregoing it can be deduced that the noise in the secondary direct current is given by:

$$\overline{i_{sec}^2} = 2eI_{pr}\delta(\kappa - \delta)Af = 2eI_{sec}(\kappa - \delta)Af \dots (8)$$

If fluctuations occur also in the primary current according to (4), (6) or (7), then these contributions towards i^2 have to be added to (8).

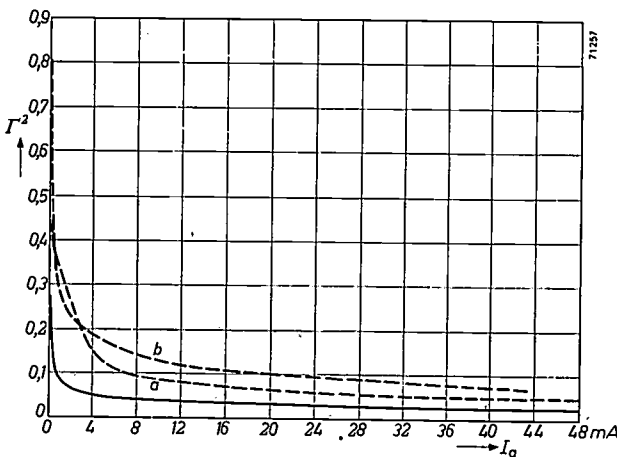


Fig. 3. Influence of the reflectivity of the anode upon the noise of a diode. Fully drawn curve: theoretical value of Γ^2 as a function of the anode current I_a for a reflection coefficient 0. Broken curves: *a* experimental values for an anode covered with fluffy soot (low reflection coefficient), *b* experimental values for a polished nickel anode (higher reflection coefficient).

Flicker effect

At frequencies lower than 10 000 c/s the anode current fluctuations in a diode often exceed the value calculated according to (6). The intensity of this additional noise appears to increase approximately in inverse proportion to the frequency. This additional noise is ascribed to local changes taking place in the cathode surface, which have rather a long characteristic duration ⁶⁾.

Ion noise

In some cases ions are emitted by the cathode or produced by the stream of electrons if gas is still present in the valve. If there is a potential minimum these slow ions may greatly influence it owing to the length of time that they remain near it. The result is an increase of the noise at frequencies lower than the reciprocal transit time, i.e. usually at wavelengths greater than a few metres.

Quantum noise

At very high frequencies another type of noise may arise, connected directly with the quantum character of the electromagnetic radiation ⁷⁾. In a cavity resonator filled with electromagnetic radiation which is in thermodynamic equilibrium with the surroundings, a number of quanta pass a given cross section every second. This number fluctuates statistically. A radiation quantum in the order of 1.3×10^{-4} eV corresponds to radiation of a wavelength of 1 cm. This means that in a cavity resonator with an alternating field of this wavelength the energy, and thus the velocity of an electron, can change only discontinuously, in steps corresponding to 1.3×10^{-4} eV. It has not yet been possible to detect the noise which must result from this.

Partition noise

In a valve having several electrodes there will be fluctuations in the distribution of electrons among the electrodes. When a certain electrode takes up on an average the proportion p of the total current I then, according to the rule of probability, the fluctuations in the current flowing towards that electrode are given by:

$$\overline{i_p^2} = 2ep(1-p)IAf \dots (9)$$

Here again, any fluctuations already present in I have to be added to those given by (9).

Electronic noise of a diode at very high frequencies

As already remarked, in the range of very high frequencies the description of noise phenomena is made very complicated by the fact that the transit times of the electrons in the valve (in the order of 10^{-10} to 10^{-8} sec) are then no longer small compared with the cycle of the alternating field. It can be shown that both the width and the waveform of the current pulses produced by individual electrons passing through the valve then begin to play a part. There is a change in the effect of the causes of noise already mentioned and other causes arise, as will be seen from the following.

⁶⁾ J. G. van Wijngaarden, Thesis "Vrije Universiteit" Amsterdam, 1951.
⁷⁾ D. K. C. MacDonald and R. Kompfner, Proc. Inst. Radio Engrs 37, 1424-1426, 1949; 38, 304, 1950.

Current pulses in an electronic valve

An electron travelling towards the anode causes a current pulse in the anode circuit owing to the induction of charges on the electrodes (cathode and anode). The shape of this pulse (its variation as a function of time) depends upon the kinetic state of the electron. Let us consider, with reference to *fig. 4*, an electron moving between two flat metal

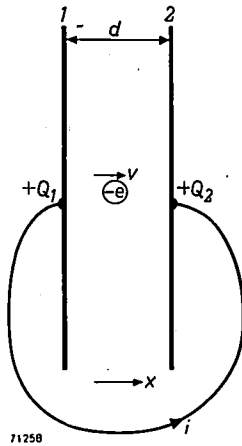


Fig. 4. Illustration of the current induced in the external lead.

plates 1 and 2 kept at the same potential by a conducting connection. The charge induced by the electron on plate 1, Q_1 , and that induced on plate 2, Q_2 , will depend only upon the distance x between the electron and plate 1, assuming that the plates are so large that fringing effects may be ignored. The charges induced may then be calculated by imagining the charge of the electron, e , as being spread out over a plane parallel to the electrodes. Denoting the distance between the electrodes by d , then:

$$Q_1 = \frac{d-x}{d} e,$$

$$Q_2 = \frac{x}{d} e.$$

Thus the current i flowing in the external circuit is:

$$i = \frac{dQ_2}{dt} = \frac{e}{d} \frac{dx}{dt}.$$

The instantaneous value of the current therefore depends only upon the velocity of the electron, and not upon its position. Hence a current will flow in the external circuit even before the electron has reached plate 2; this is an induced current. In the valve one may speak of a convection current, the instantaneous value of which is related to the position of the electrons: this instantaneous value is equal to the net amount of charge flowing per unit time through a given cross section of the valve.

At low frequencies the convection current is equal to the current in the external circuit. At frequencies which are in the order of the reciprocal transit time of the electrons account has to be taken of the induced current, and thus the convection current at any instant is only indirectly related to the current flowing at that instant in the external circuit. The convection current cannot then be determined by measurement.

At low frequencies the current pulse generated in the external circuit by an individual electron can be considered infinitely narrow. *Fig. 5a* shows diagrammatically how the current produced in the external circuit by all the electrons passing through the valve can then be imagined as comprising a series of infinitely narrow pulses (assuming that we have a saturated diode with negligible space charge). A Fourier analysis of such a current shows that all frequencies are equally represented in it (*fig. 5b*). Therefore, at low frequencies the electronic noise of a valve may be regarded as independent of frequency (formula (4)).

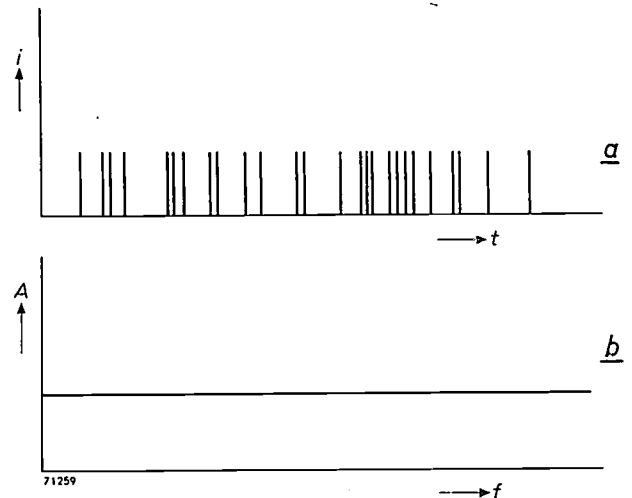


Fig. 5. (a) At low frequencies the current generated in the external circuit of a diode by the individual electrons can be regarded as a series of infinitely narrow pulses. (b) The Fourier spectrum of such a series has an amplitude A which is independent of the frequency.

At high frequencies it must be taken into account that in reality the current pulses are not infinitely narrow but of finite width and are erratic in shape. A series of such pulses is illustrated in *fig. 6a*. (It should be recognized that actually these pulses follow each other in such rapid succession that they almost entirely overlap, so the illustration is not true to actual conditions.) A Fourier analysis of a pulse of this shape shows that in the spectrum there is an almost constant part extending to frequencies in the order of the reciprocal duration of the pulse, i.e. the reciprocal transit time of the

electron (fig. 6b). At higher frequencies the noise caused by the electrons decreases, as already mentioned in the discussion on the saturated diode as a noise standard.

The shape of the current pulses in a diode will now be considered in greater detail.

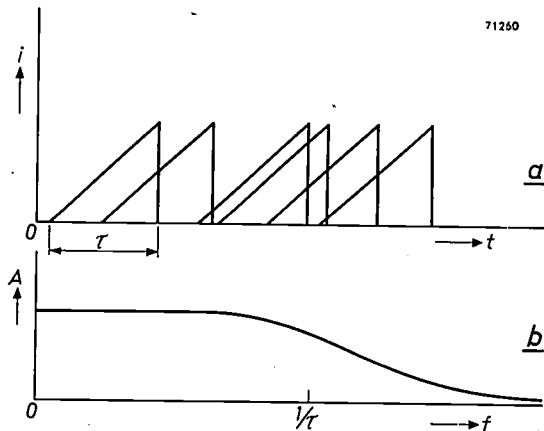


Fig. 6. (a) At very high frequencies the individual current pulses in the external circuit of a valve can no longer be treated as infinitely narrow pulses; their width is in the order of the reciprocal transit time τ of the electrons. In this diagram the pulses have been fairly well separated, but actually they follow each other in such a rapid succession as to overlap almost completely. (b) The amplitude of the Fourier spectrum of this series of pulses begins to decrease at frequencies in the order of the reciprocal transit time.

Current pulses in a diode at very high frequencies

Again we shall consider the case of a diode with planar parallel electrodes. If the anode voltage is so high and the emission current so small (low cathode temperature) that the velocity with which they leave the cathode and the space charge may be ignored, the electrons pass from the cathode to the anode with a uniform acceleration. The shape of the current pulse generated by an electron in the short-circuited external circuit is then triangular (fig. 7a). If there is space charge, but the resultant potential minimum lies practically on the cathode surface (as may be the case when the anode voltage is still rather high), then to a good approximation the potential in the valve varies, according to Child⁸⁾, as:

$$V(x) \sim x^{3/2}.$$

From this it can be deduced that the velocity of an electron at a moment t after leaving the cathode is given by an expression of the form:

$$v(t) \sim t^2. \dots \dots \dots (10)$$

The shape of the current pulses is then parabolic,

as represented in fig. 7b. In these figures τ stands for the total transit time of the electron.

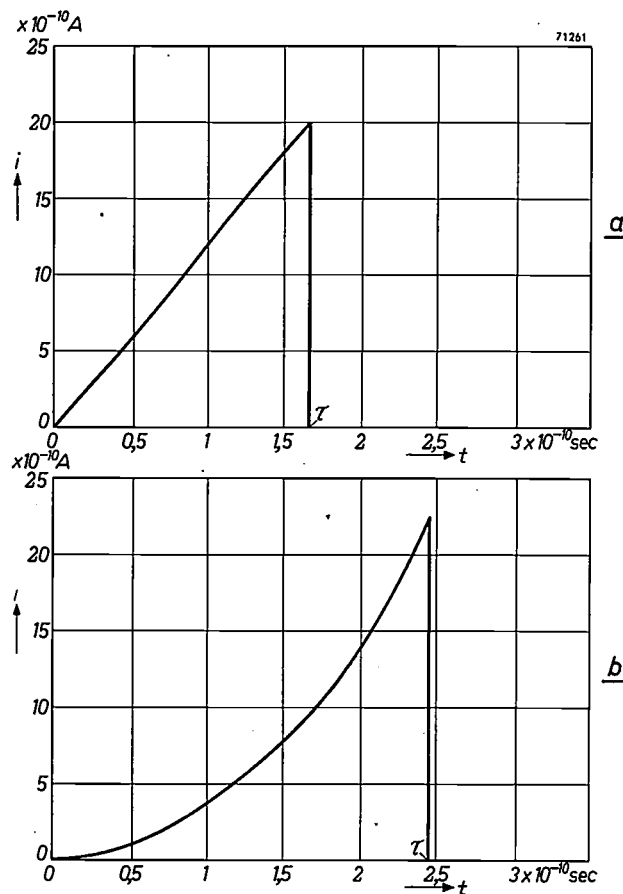


Fig. 7. Shape of the current pulse (variation of the current i as a function of the time t) generated in the external circuit of a flat diode, by an electron passing from cathode to anode: (a) for a diode without space charge, (b) for a diode answering Child's law (formula (10)); τ stands for the transit time of the electron. This is only a schematic diagram.

If the space charge is so great that the potential minimum no longer lies on the cathode but somewhere between the cathode and the anode, then various shapes of current pulses are possible. There are then three distinct categories of electrons (see fig. 8):

1) α -electrons — these are electrons which cannot pass the potential minimum and which therefore return to the cathode. They usually form the major part of all the electrons, so that the number of these α -electrons starting per second can be taken to be practically equal to the total stream of electrons leaving the cathode. They are, in fact, often referred to as total-emission electrons, and the phenomena to which they give rise are called total-emission phenomena. The current pulses of α -electrons are roughly of the shape depicted in fig. 9a. After leaving the cathode with a certain initial velocity at the moment $t = 0$, an electron is retarded and at the moment τ_a turned

⁸⁾ D. C. Child, Phys. Rev. 32, 492-511, 1911; I. Langmuir, Phys. Rev. 2, 450-486, 1913.

back. It is then accelerated in the reverse direction until it reaches the cathode again at the moment $2\tau_\alpha$. In the vicinity of τ_α part of the slope of the current pulse is less steep, and the closer the electron gets to the potential minimum before turning

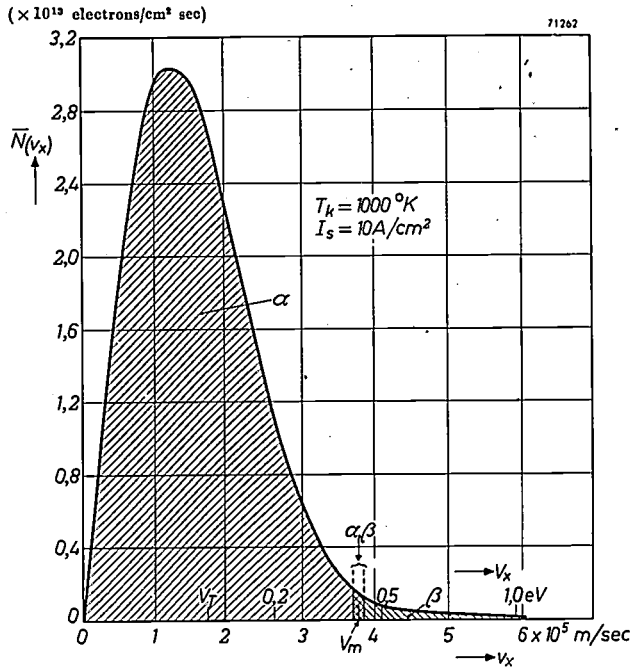


Fig. 8. Velocity distribution, according to formula (5), of electrons leaving the cathode. The mean velocity corresponds to V_T (expressed in electron-volts). The numbers given are practical figures. The majority of the electrons do not reach the anode (α -electrons); those which do reach it are much fewer in number (β -electrons). The α - β -electrons have velocities within the transitional zone between the α - and the β -electrons — round about the velocity v_m , which is related to the depth of the potential minimum V_m as $v_m = \sqrt{2eV_m/m}$. The behaviour of these electrons is greatly affected by fluctuations in the depth of the potential minimum. v_x is the forward electron velocity. The lower horizontal scale gives v_x in metres/sec, the upper one the corresponding energy V_x in electron-volts.

back, the longer this shallow part becomes. If there were no fluctuations in the minimum, the electrons which just managed to reach it would have an infinitely long transit time.

In addition to the current pulses already described also secondary current pulses occur, owing to a temporary lowering of the potential minimum during the time that an α -electron is close to it. As a result there is a temporary increase in the number of electrons forced to return to the cathode. This phenomenon will be referred to again later. 2) β -electrons — electrons leaving the cathode at such a velocity as to be able to pass the potential minimum. These electrons are therefore responsible for the anode current I_a . The current pulses resulting from the β -electrons have the shape depicted in fig. 9b. After leaving the cathode such an

electron is decelerated, but once it has passed the potential minimum (at the moment $\tau_{\beta m}$) it is accelerated until it reaches the anode (at the moment $\tau_{\beta m} + \tau_{\beta a}$). Since these electrons, too, temporarily deepen the potential minimum, they will also give rise to secondary current pulses.

3) α - β -electrons — these are the electrons leaving the cathode with velocities round about the boundary velocity between α - and β -electrons, which corresponds to the potential V_m in fig. 8. These electrons are highly sensitive to small variations in the depth of the potential minimum, such as may arise, for instance, from slight variations in the anode voltage. Though few in number, they nevertheless have a considerable influence upon the properties of the valve; for example, they determine the increase or decrease of anode current when the anode voltage varies. It may, therefore be said that the α - β -electrons are responsible for the static (low-frequency) internal resistance of the diode.

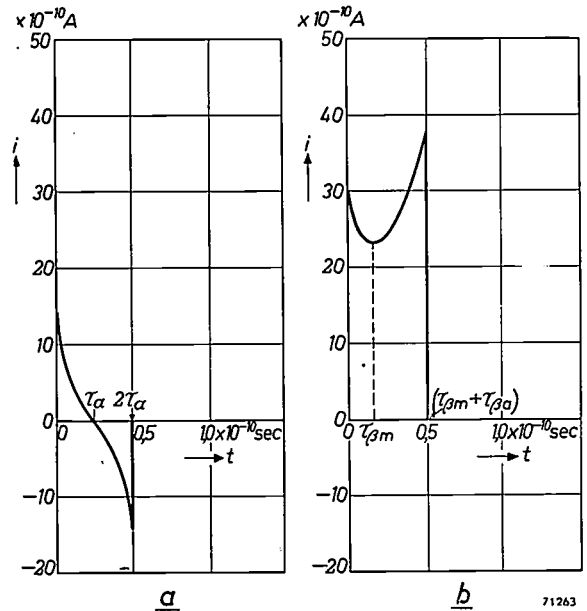


Fig. 9. (a) Shape of the current pulse generated by an α -electron in the external circuit of a diode with space charge. (b) The same for a pulse generated by a β -electron. The symbols are explained in the text.

Fluctuations in the potential minimum; secondary current pulses

When there is one α -electron in transit in addition to the average number of electrons, and particularly when that electron stays in the vicinity of the potential minimum (round about τ_α in fig. 9a), the depth of that minimum changes, as indicated in fig. 10⁹). Where, in the following, mention is made of the influence of one single electron, it

should be taken as the average influence for a group of electrons all under practically the same

conditions. There are in fact so many electrons leaving the cathode per second that they may be considered as being in groups within which there are only negligible differences in initial velocity and the instant of emission. The movement of such a group therefore influences the anode current: more α - β -electrons turn back. Its effect is equivalent to two secondary current pulses in the external circuit, firstly that caused by a temporary shortage of traversing electrons (i.e. by "holes" starting in the minimum and moving towards the anode in a more than linearly accelerating field, with transit time $\tau_{\alpha\beta a}$), and secondly that caused by an equally large surplus of returning electrons (moving towards the cathode in a more than linearly accelerating field, with transit time $\tau_{\alpha\beta c}$). These two current pulses are represented diagrammatically by curves *I* and *II* respectively in fig. 11a. Both the "holes" and the surplus electrons stay in the vicinity of the potential minimum a relatively long time, so that $\tau_{\alpha\beta a}$ and $\tau_{\alpha\beta c}$ are much greater than $\tau_{\beta a}$ and τ_a .

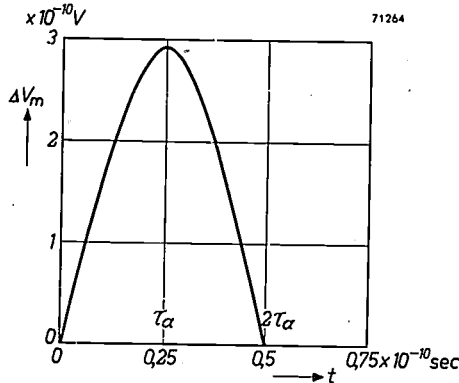


Fig. 10. An α -electron moving somewhere between anode and cathode (transit time $2\tau_a$) causes a gradual change ΔV_m to take place in the depth of the potential minimum.

9) Except for the sign, a shortage of emitted electrons causes the same phenomena as a surplus. Since it is the fluctuations that are being considered here, the sign is immaterial. In addition to the fluctuations in depth of the potential minimum, fluctuations also occur in its position. The latter, however, produce effects of a higher order, which can be left out of consideration here.

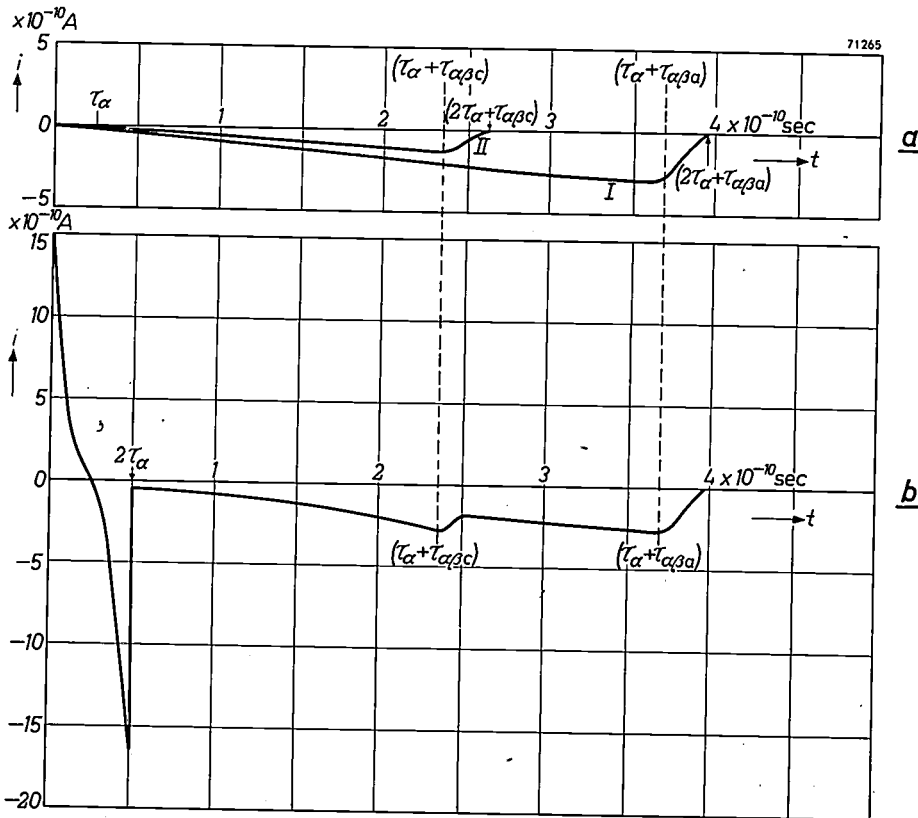


Fig. 11. (a) As a result of the change in depth of the potential minimum, fewer α - β -electrons reach the anode. This may be represented (*I*) by a stream of „holes” moving from the potential minimum towards the anode (with transit time $\tau_{\alpha\beta a} (> \tau_a)$), commencing perceptibly at the moment τ_a and rapidly decreasing after a time $\tau_a + \tau_{\alpha\beta a}$ (with decay time τ_a); and (*II*) a stream of additional electrons returning to the cathode with a transit time $\tau_{\alpha\beta c} (> \tau_a)$. This current also begins to become of importance after τ_a , and continues for an interval of time $2\tau_a + \tau_{\alpha\beta c}$. (b) The composite (total) current pulse produced by an α -electron in the external circuit. This is practically identical to the primary current pulse of fig. 9a, but has a small surplus on the negative side. Here this surplus is greatly exaggerated.

Since but few α -electrons penetrate to close to the minimum, the amplitude of the pulses I and II is only small compared with the primary pulse of the α -electrons, especially when the anode voltage is not too low (little space charge). The total current pulse of the α -electrons is then practically equal to the primary pulse; it is represented in fig. 11b. The resulting area below the curve is practically equal to zero. This means that at low frequencies the amplitude of the Fourier spectrum is almost zero. There is, however, a small excess below the axis, which represents the very weak total-emission noise at low frequencies. From fig. 11b it can be seen that for high frequencies the pulse has an A.C. character, a wave with a cycle of about $2\tau_a$ predominating. The Fourier spectrum of this pulse shows large amplitudes at frequencies in the order of $1/(2\tau_a)$, which is mostly in the centimetric wave range. Relatively large alternating currents with such frequencies therefore arise in the external circuit and cause noise, which is called the high-frequency total-emission noise¹⁰⁾.

¹⁰⁾ A. van der Ziel and A. Versnel, Philips Res. Rep. 3, 13-23, 1948; J. J. Freeman, J. Res. Bur. Standards 42, 75-88, 1949.

When, during their transit time, electrons returning to the cathode ^{FIND THEMSELVES} stay in an external alternating field having a period of oscillation that is no longer very ^{LARGE} small compared with the transit time, their movements give rise to periodical variations in the rhythm of that field. As a result, alternating currents of the same periodicity are induced in the external circuit. These currents are, as a rule, phase-shifted with respect to the external alternating field and constitute a load on the voltage source producing that field. Owing to the phase shift the load can be regarded as a complex admittance, the total-emission admittance, the real part of which is called the total-emission conductance.

This conductance may be considered as being an equivalent noise resistance, which is responsible for the total-emission noise. It is then found that this resistance has an equivalent temperature equal to the cathode temperature — not surprisingly considering that the α -electrons are practically in thermal equilibrium with the cathode.

A β -electron also brings about a change in the depth of the potential minimum — a change much

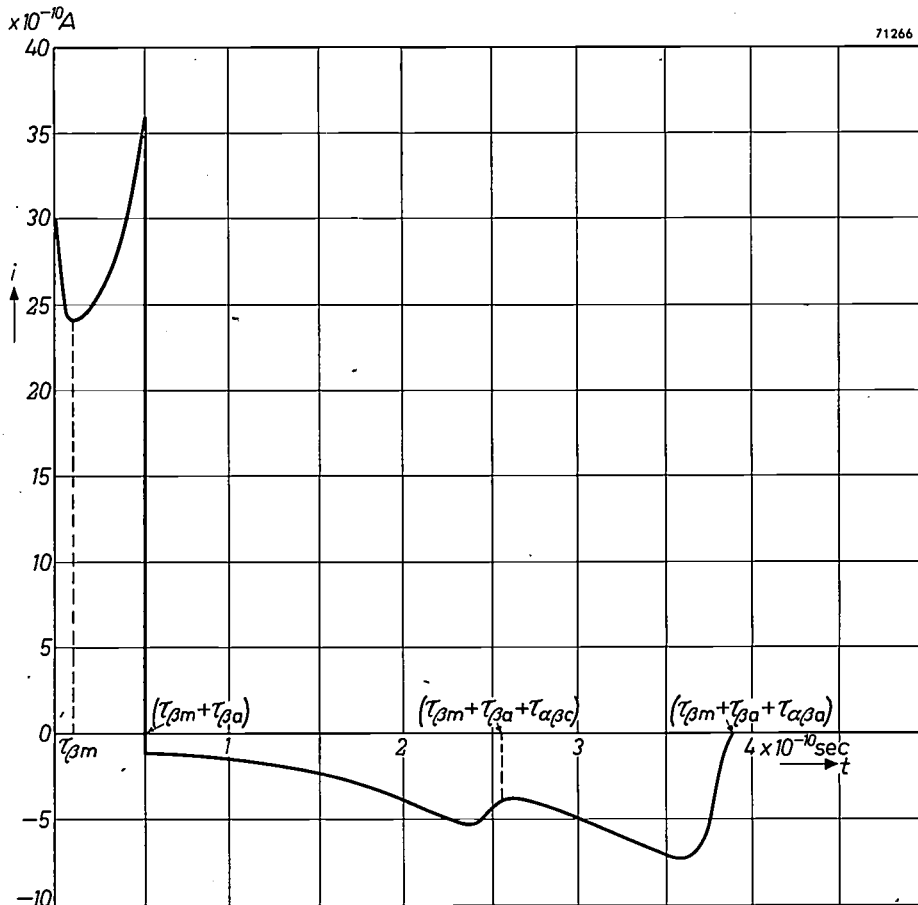


Fig. 12. Shape of the composite current pulse caused by a β -electron. Here the secondary current pulses are much stronger than those of the α -electrons. The index β relates to the primary pulse, the index $\alpha\beta$ to the secondary pulses.

greater than that caused by an α -electron, because on an average β -electrons stay much longer in the vicinity of the minimum. The secondary current pulses produced by the β -electrons are practically identical in shape with those of fig. 11a, but of much greater amplitudes, in accordance with the above comments. The shape of the composite β -current pulse — primary and secondary together — is represented in fig. 12. Averaged over various initial velocities, the area of the curve above the axis is practically equal to that below the axis, as it was for α -electrons. Here this is due to space charge, but the effect is similar: at low frequencies the β -electrons cause practically no noise. The Fourier spectrum of the β -pulses contains, however, strong

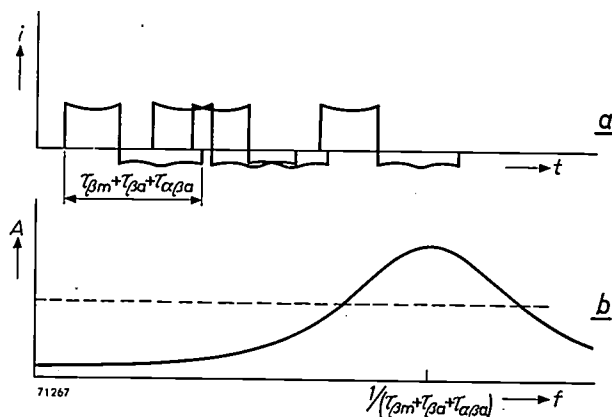


Fig. 13. As in fig. 6, but for a diode with space-charge suppression. (a) Shape of the current pulses (see fig. 12), (b) noise spectrum. At low frequencies the noise is much weaker than the non-suppressed noise in fig. 6b (denoted by the broken horizontal line), but at frequencies in the order of the reciprocal transit time it may increase to about twice that value. At still higher frequencies it decreases again.

components near the high frequency $f = 1/(\tau_{\beta m} + \tau_{\beta a} + \tau_{\alpha \beta a})$, so that at that frequency the β -electrons do yield an appreciable contribution towards the noise. In fig. 13a a series of overlapping (β -) pulses have been drawn, similar to fig. 6a, whilst fig. 13b represents the corresponding noise spectrum. Owing to noise suppression, the intensity at low frequencies is less than that according to fig. 6b. At frequencies in the order of the reciprocal transit time, however, the intensity may be about twice that of the non-suppressed noise of fig. 6b, because of the special waveform of the pulses. At still higher frequencies the intensity again approaches zero.

The conclusion to be drawn from all this is that, owing to the influence of the electrons upon the potential minimum, electronic valves will show more noise at very high frequencies than at intermediate ones. This is borne out by experimental facts, as is illustrated in fig. 14.

Transit-time functions in a diode

In the foregoing we have investigated the shape of the current pulses induced in the external circuit by the movement of electrons inside the valve. Fourier analysis of these current pulses revealed that at frequencies in the order of the reciprocal transit-times of the electrons inside the valve, the noise in the external circuit is stronger than at lower frequencies. Nothing can be said against this method physically, but it does not lend itself well to quantitative calculations, and for that reason an approximative method is often employed.

First a Fourier analysis is made of the convection-noise current i_{con} caused by the electrons leaving the cathode. From the spectrum a component $i_{con}(f)$ of a certain frequency is then chosen and its relationship to the noise-current component $i(f)$ of that frequency in the (short-circuited) external circuit investigated. This relation is given by a transit-time function Ψ_1 defined as:

$$i(f) = i_{con}(f)\Psi_1.$$

Ψ_1 is a complex function of the argument jft , where τ stands for the average transit time from cathode to anode.

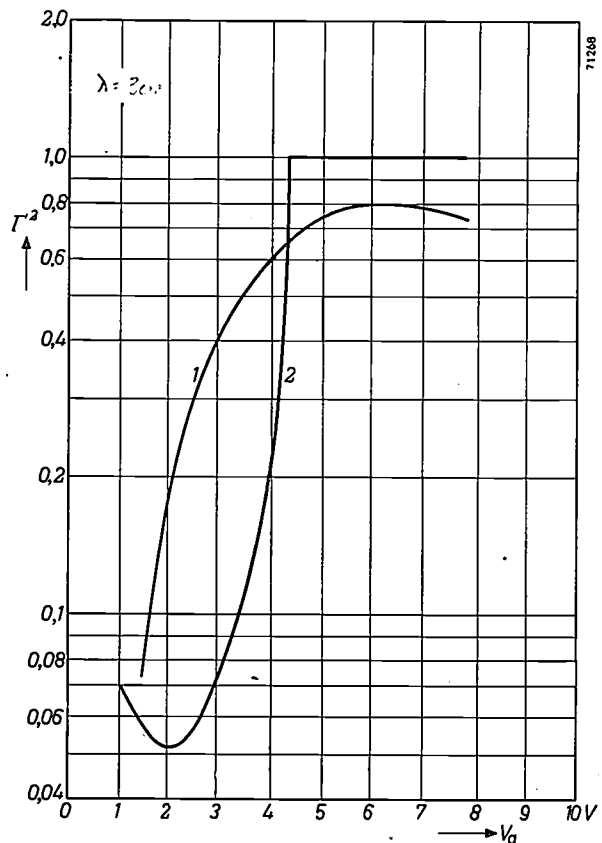


Fig. 14. The experimental value of Γ^2 for an experimental diode with L-cathode at a wavelength of 3 cm (curve 1) and at moderate anode voltages is much higher than the theoretical value at low frequencies (curve 2).

When calculating Ψ_1 it is mathematically very difficult, if not impossible, to take into account the interaction between the electrons near the potential minimum. But after certain simplifications, among them the assumption that the initial velocities of the electrons may be ignored, it is possible to compute Ψ_1 ¹¹⁾.

An alternating signal voltage across the valve will influence the depth of the potential minimum, and set up an alternating convection current I_{con} . This produces an alternating signal current I in the external circuit, which can also be expressed in terms of I_{con} with the aid of a transit-time function. This function, Ψ_2 , differs from Ψ_1 in that the electrons are here moving in an alternating potential field whose phase bears a certain relationship to the phase of I_{con} . The noise currents, on the other hand, bear no relationship at all to the signal voltage. The transit-time function Ψ_2 has been calculated with the aid of the approximation mentioned above. For a better approximation the static depth and the position of the potential minimum have sometimes been taken into account, but the high-frequency fluctuations in the potential minimum have hitherto always been ignored¹²⁾.

Notwithstanding these approximations, transit-time functions have proved to be very useful in practice, but when applying them one should always take the results *cum grano salis*.

What has been said here about the diode applies also to valves with more electrodes, such as the

¹¹⁾ C. J. Bakker and G. de Vries, *Physica* 2, 683-697, 1935.
¹²⁾ A. van der Ziel, *Wireless Eng.* 28, 226-227, 1951.

triode, but then the noise phenomena are influenced to a much greater extent by the circuit in which the valve is incorporated. The mutual influence of the currents in the input and output leads varies according to the nature of the circuit. In the case of a triode it appears to be possible for the circuit to be so arranged that the noise is partly compensated by this mutual influence, but such compensation is out of the question for a diode, since it has only one external lead. The noise of a triode will be discussed in another article to follow. It will be seen then that, though the complexity of the problem makes it necessary to employ the method of approximation with the aid of transit-time functions, it is possible nevertheless to draw very important conclusions for practical purposes.

Summary. Following upon a review of the various causes of noise inside and outside a radio receiver, a closer study is made of the noise resulting from the movement of electrons in a diode (electronic noise). At not too high frequencies (lower than 10^7 c/s) there are a number of causes of noise, amongst others the shot effect, which are briefly discussed. The influence of possible space charge upon the potential distribution in the valve, and thus upon noise is investigated. At frequencies round about 10^8 c/s and higher, where the transit times of the electrons can no longer be regarded as being infinitely short, other noise phenomena arise. These are related to the finite width and special shape of the current pulses generated by individual electrons in the external valve leads. This causes, among others, total-emission noise. At very high frequencies signal currents also are subject to the influence of the finite width of the current pulses of which they are composed, and one of the results of this is an increase in the input conductance (total-emission conductance). Consideration is given to the different behaviour of noise currents and signal currents in the valve. Both these kinds of current can be approximately determined by means of (different) transit time functions.

AN EXPERIMENTAL X-RAY APPARATUS WITH MIDGET X-RAY TUBE

by P. J. M. BOTDEN, B. COMBÉE and J. HOUTMAN.

621.386.1 : 615.849 : 620.179

The small, mass-produced and inexpensive high-tension generator of a projection television receiver has been found to be a most suitable voltage supply unit for a very small X-ray tube. The present article shows how this fact was made use of when the "KT" apparatus was developed. The X-ray tube of this apparatus is not only remarkable for its small size, but also for a directional effect in the emitted radiation, favourable to therapeutic application, which effect is not observed in normal X-ray tubes.

Introduction

The X-ray tube which we will discuss in the present paper is probably the smallest ever designed. It is 45 mm long and 14 mm in diameter, including the earthed jacket. A power of 2.5 watts can be dissipated continuously at the anode, the maximum tube voltage is 25 kV, and the maximum tube current 200 μ A. To complete this brief description, *fig. 1* shows a photograph of the tube (which has

therapeutic applications of X-rays it can be most useful to the physician to have an X-ray source at his disposal, of such small dimensions that it can easily be held and positioned by two fingers, and which can be inserted into small cavities of the body if necessary. Sometimes radiographs of all sorts of objects have to be made for research or teaching purposes, where a very soft radiation is required, an

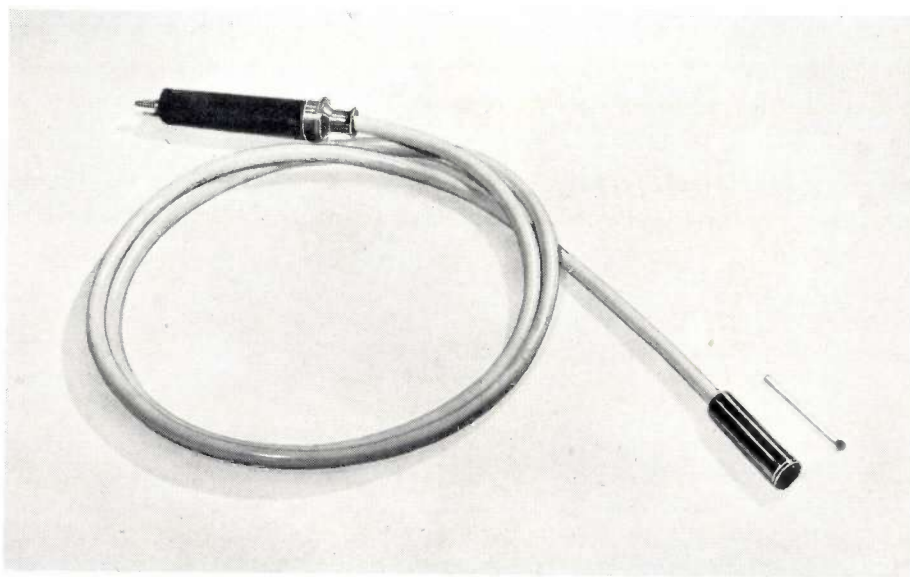


Fig. 1. Midget X-ray tube (KT tube) with cable and plug. A match is shown by the side of the X-ray tube (to the right) for comparison.

not yet been brought into regular production), with the cable connecting it to the high-tension source. The larger tubular element at the left end of the cable is not the X-ray tube, but the plug.

A few remarks will show that the design of this midget X-ray tube (and the accessory high-tension source) was not merely inspired by an ambition to design the smallest X-ray tube of the world. In certain

extremely small power is sufficient and a high price for the apparatus is undesirable. This defines in general terms the possible application of the above characterised "KT" tube and "KT" apparatus (from the Dutch for "smallest therapy", viz. "kleinste therapie"). We shall discuss below the special features of this X-ray source, and those it lacks; it is convenient to compare it with the CT tube

(CT from Contact Therapy), described previously in this review, which, of all X-ray tubes is most similar to the tube under discussion ¹⁾²⁾.

We shall conclude the present paper with a few further observations on the possibilities of application.

Construction of the X-ray tube

In *fig. 2* a schematic sectional view of the tube is given. The simplicity and compactness of design shown by the figure are chiefly due to the fact that the X-rays leave the tube via the anode. For this

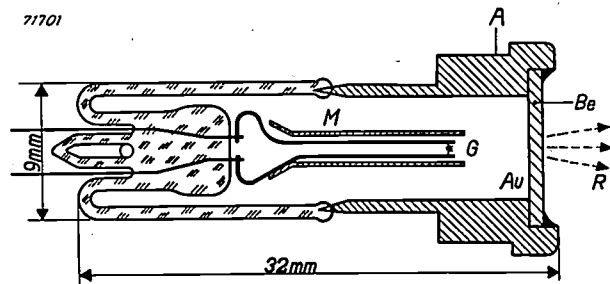


Fig. 2. Schematic cross-section of the KT tube. A anode can, Be beryllium plate, Au gold layer, G filament, M metal cylinder, R effective X-ray beam.

purpose the anode consists of a very thin layer of gold deposited on a vacuum-tight plate of beryllium. Due to the very high atomic number of gold (79) a high efficiency of X-ray production is obtained from the electrons slowed down in it; owing to the very low atomic number of beryllium (4) this will transmit the soft radiation excited at 25 kV with negligible attenuation. The gold layer has to be very thin in order to restrict the proper absorption of the excited X-radiation to a minimum; on the other hand it should not be too thin, since all electrons passing through the gold layer are naturally lost for the generation of X-rays.

The "inherent filter" of the X-ray tube, i.e. the filter which the X-rays, produced by the tube designed in the manner given above, must pass before their practical application, is equivalent to about 1.5 mm beryllium. The importance of this very low inherent filter will be discussed presently.

The beryllium plate is soldered in a metal bush which functions as tube wall for this end of the tube ("anode can"). This metal anode assembly is earthed and the cathode is connected to high voltage. The cathode comprises a filament, consisting of a tung-

sten coil of a few turns, and a small metal cylinder enclosing the filament and its supporting poles. This cylinder has a focusing effect on the electrons emitted by the filament. Moreover, it simplifies the problems of the insulation between the cathode and the closely adjacent earthed jacket: without the cylinder the electric field strengths at the thin lead-in poles would be much higher than they now are on the cylinder.

The filament heating power can amount to about 1 watt at the highest (1.4 V, 0.6 A). The coiled filament will then give a saturation emission of 200 μ A.

The focus and the radiation obtained

A description of an X-ray tube is not complete without some remarks regarding the configuration and the loading of the focal spot. In this case it is most appropriate to begin with some general observations, which will also explain the choice of the constructional principle adopted.

Soft radiation is suitable for the therapeutic irradiation of very thin layers of tissue on or at the surface of the body. The underlying healthy tissue receives only relatively small doses in this case and is consequently very little affected (small depth penetration). Owing to the low voltage and the very small inherent filter, the radiation of the tube is extremely soft; the half value thickness of the radiation is only 0.035 mm of aluminium or 0.5 mm of tissue at a tube voltage of 25 kV, and at 10 kV it is even 0.02 mm of Al or 0.3 mm of tissue. In this respect the radiation is similar to that of the present CT tube with a window of mica and beryllium (see the article referred to in note ²⁾).

A second means of obtaining a very small depth penetration is to place the focus of the X-ray tube at a very short distance from the skin. This was extensively commented on in the article referred to ¹⁾. This same article further stated that the smallest focus-skin distance, and therefore the smallest relative depth doses can be realised if the anode itself serves as exit window for the X-rays. This design principle also leads to a relatively simple tube construction. It is generally considered to be a drawback to this principle of design that the rays passing obliquely through the anode are attenuated more than those passing through perpendicularly, resulting in a strong decrease of the dosage rate from the centre to the edges of the irradiated area. This makes it difficult to apply an exact dosage of radiation. For this reason a different design was chosen for the CT tube, as may be seen in the article mentioned ¹⁾.

¹⁾ H. A. G. Hazen, J. M. Ledebuer and J. H. van der Tuuk, An X-ray apparatus for contact therapy, Philips Techn. Rev. 8, 8-15, 1946.

²⁾ B. Combée and P. J. M. Botden, Special X-ray tubes, Philips Techn. Rev. 13, 71-80, 1951 (No. 3).

We have, nevertheless, already observed that the design with anode functioning as window was chosen for the KT tube. This does not imply, however, that the objection referred to above was ignored; surprisingly enough this objection does not hold at all in this case! If the radiation intensity of the KT tube is measured in various directions³⁾, a diagram as given in *fig. 3* results. The intensity by

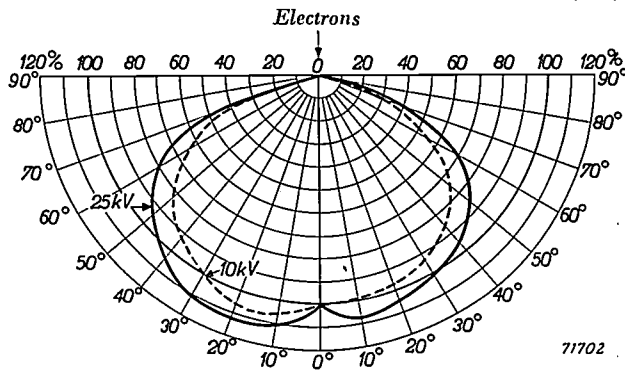


Fig. 3. Relative dosage rate of the KT tube, measured in different directions from the axis, at a distance of 150 mm from the focus, for two different tube voltages. In the axial direction (0°) the dosage rate for both curves is made equal to 100. The asymmetry may be due to a slight inhomogeneity of the gold layer. (These measurements and those of *fig. 4* were made by Dr. W. J. Oosterkamp and J. Proper.)

no means decreases with increasing deviation from the axial direction, but even increases initially; at an angle of 20° to the axis it is at its maximum, and the subsequent decrease is still rather gradual up to angles of about 50°. There is thus no question of a strong decrease in the dosage rate over the irradiated area from the centre to the edges. On the contrary, a very uniform dosage distribution is found for not too large, flat areas, even more uniform than with the spherical radiation diagram of normal X-ray tubes, since in our case the larger distance from the focal spot to the off-axis parts of the field is neutralised by the initial increase in intensity shown in *fig. 3*.

The phenomenon illustrated by *fig. 3* has been known for a long time, and can be explained by a consideration of the mechanism whereby X-rays are produced in the anode. It is not observed in normal X-ray tubes because of the rather thick anode these tubes always contain. In the present case only the very thin foil of gold functions as an anode, and as a matter of fact Kulenkampff was able to demonstrate the directional effect as early as in 1928 by the use of specially designed tubes with a very thin anode⁴⁾.

³⁾ For these measurements see: W. J. Oosterkamp and J. Proper, *Acta Radiologica* 37, 33-43, 1952 (No. 1). Therein is given a full description of the method whereby such soft and heterogeneous X-rays may be measured.

⁴⁾ H. Kulenkampff, *Ann. Physik* IV, 57, 597, 1928. For more recent investigations into the phenomenon see: G. Sesemann, *Ann. Physik* V, 40, 66, 1941; O. Blunck, *Ann. Physik* VI, 9, 373, 1951.

The electrons striking the anode with great velocity are slowed down on penetrating into the metal. According to classical theory, each electron will emit radiation, the continuous X-radiation or "Bremsstrahlung". According to this theory one can also expect the radiation to have zero intensity in the direction of movement of the electron, as also in the reverse direction, whereas the radiation will be at its maximum at a certain angle to this direction — depending on the electron velocity. In a thick anode a penetrating electron may repeatedly change its direction by small amounts, due to deflection by atomic nuclei, with very small loss of energy each time, before it passes a nucleus so closely that it is strongly decelerated and emits the X-radiation observed. The contribution to the radiation, of this electron, will in that case possess the directional dependence as mentioned above, but with its last direction of movement as axis. Since these axes, for all the electrons, will have a random distribution of direction, the total radiation can reveal little or nothing of the described directional effect.

On the other hand, if the anode is very thin, the electrons will already have passed the entire layer after relatively few encounters with atom nuclei. The emission of each contribution to the observed radiation can therefore only be preceded by few deflections; consequently the "axes" for all contributions deviate but slightly from the axis of the X-ray tube, and the total radiation will consequently also show the directional effect with but little spread.

If the anode functions as exit window, the opportunity for the natural directional effect of the radiation to be observed becomes still smaller in the case of a rather thick anode, since the earlier mentioned absorption effect (intensity decrease with increasing angle), especially important in the case of thick anodes, is superimposed on this directional effect. *Fig. 3* clearly shows the increasing importance of this effect, even with the very thin anode of the KT tube, as the tube voltage is decreased (softer radiation).

Owing to the constructional principle adopted, it is possible with the KT tube to reduce the focus-skin distance to 1 mm if so desired, viz. by pressing the anode into the surface to be irradiated. Enormous dosage rates can be realised with only a very small power, if the distance between focus and object is so small. With the above-mentioned maximum permissible anode dissipation of 2.5 W and a voltage of 25 kV, the KT tube will give 17 000 röntgen units per minute at a distance of 1 cm from the focus; at 10 kV and 2 W dissipation (the tube current is 200 μ A maximum), it will still give 7000 r/min! These are measured values (*fig. 4*); a separately determined correction factor was applied for the absorption of the soft radiation in the window of the dosimeter (see the article referred to in note³⁾). At a focus distance of 1 mm even much larger dosage rates, of some hundred thousands r/min, are obtained. Naturally, if the focus distance is not very small, e.g. 20 mm, which can also be reached with the CT tube, the KT tube is, as far as dosage rates are concerned, by far the inferior of the CT tube, which was designed for a maximum power of 100 W and voltages up to 50 kV.

In normal X-ray therapy tubes the focus is made as small as possible in order to obtain a sharp boundary to the field to be irradiated (small penumbra width of diagram edges); in the KT tube, on the

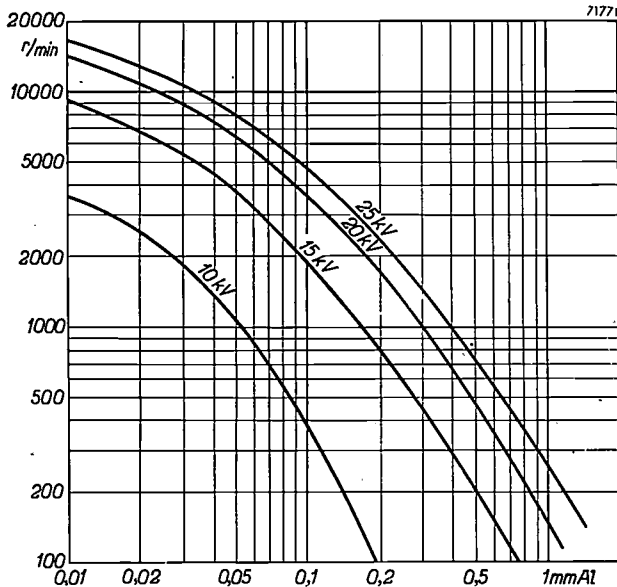


Fig. 4. Absorption characteristic of the radiation of the KT tube. The dosage rate obtained at 1 cm distance from the focus when the radiation is weakened by an aluminium filter of varying thickness is plotted for four values of the tube voltage. The minimum filter, the inherent filter of the tube, is equivalent to about 1.5 mm beryllium or 0.03 mm aluminium. For this case the graph gives a value of 17 000 r/min at 25 kV and 3600 r/min at 10 kV. (The figures along the horizontal axis are not correct: a constant value of 0.02 is to be added to all of them.) All curves were measured at a tube current of 100 μ A; the 25 kV curve, therefore, corresponds to a load of 2.5 W, the 10 kV curve to 1 W. The measurements were made at a distance of 2.5 cm from the focus and recalculated to a distance of 1 cm by use of the inverse square law: a correction was made for the smaller air absorption at 1 cm distance as against 2.5 cm distance from the focus.

contrary, the electrode configuration was so designed as to spread the focus over the entire anode surface (diameter 6.5 mm). Owing to the rather large focal dimensions the loading of 2.5 W is permissible and the above-mentioned very large dosage rates are obtained. It is true that part of the gain with regard to the depth penetration, obtained by the selected principle

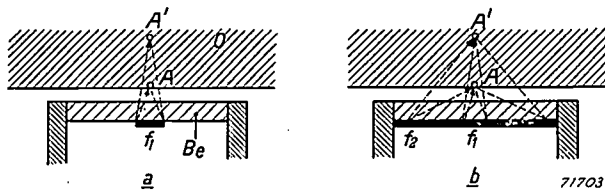


Fig. 5. (a) If the focus (f_1) is small, the dosage rate decreases as the inverse square of the distance. The dose thus rapidly decreases from A to e.g. A'. (b) If the focus is large, the dosage rate decreases more slowly than as the inverse square of the distance. The decrease is intensified, however, by the extra absorption of oblique rays in the object O: point A' profits less from the largeness of the focus than point A, since the rays originating from f_2 have to travel a longer distance in the tissue than those originating from f_1 .

of construction, is sacrificed if the focus is so large, for if the focus is not small in proportion to the focus-skin distance, the dosage will, for geometrical reasons, not follow the inverse square law, but will decrease much more slowly with increasing distance. Thus, the dosage rate at 1 mm distance will not be 100 times larger but, according to our computation, only 25 times as large as at 1 cm distance. Still, one can see from these figures that the distance effect is still considerable. Moreover, it is enhanced by the effect of the absorption of the rays in the tissue: points at some distance below the skin surface profit less by the largeness of the focus, therefore, than do points on the skin surface (see fig. 5). The resulting decrease of the dosage with depth is shown in fig. 6. It will be seen, consequently, that only a thin layer of skin of a few tenths of one millimetre is effectively irradiated.

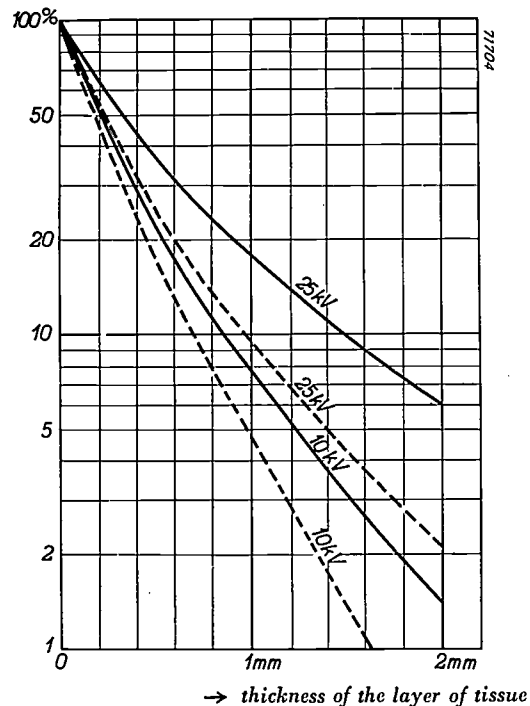


Fig. 6. Dosage rate of the KT tube computed for various depths in the irradiated body, at a focus-skin distance of 1 mm, for two different tube voltages. The dosage rate on the skin was set equal to 100 for both curves. The broken-line curves represent the dosage rate which would be obtained with a point focus.

It will be seen from this, that the lack of sharpness of the field boundary, which was mentioned above as the current objection against a broad focus, will not be of much significance in this case. For this lack of sharpness will only be appreciable in planes at some distance behind the diaphragm, in the case under discussion — the diaphragm being represented by the anode cap placed on the skin — at a certain depth below the skin where the dose is already relatively small.

If the anode is placed on or near to the skin of the patient, the focus is of nearly the same area as the irradiated skin surface. Naturally this is conducive to a still further equalisation of the dosage distribution over the irradiated surface (see above).

The above-mentioned limit of 2.5 watts to the loading of the anode is set by the temperature both of the focus and the jacket. We need not distinguish between continuous and short-duration loading in the case of the KT tube. The heat capacity of the gold foil is so small that the maximum focus temperature is always reached, even in the briefest period of loading likely to be used. The temperature of the outer face of the anode rises in proportion to the time of loading, in accordance with the curve

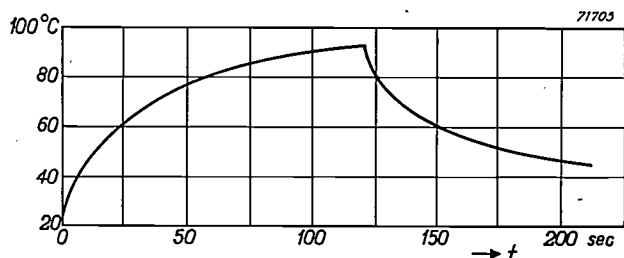


Fig. 7. Temperature time curve for the outer surface of the anode of the KT tube in free air under load. The tube was loaded to 2.5 W from $t = 0$ to $t = 120$ sec, the tube then being switched off. (If the anode is placed on the skin of the patient, the temperature will rise a little more slowly, owing to the thermal conductivity of the body.)

represented in *fig. 7*, which shows that on prolonged loading temperatures not much higher than 90°C are reached. This temperature is, of course, too high for direct contact with the skin of the patient, so that irradiation in this manner is limited to 10 or 20 seconds — which will, however, in general be amply sufficient for the application of the necessary dose!

For irradiation at a certain, albeit small, distance (in which case e.g. a small localizer is fitted to the tube) the anode temperature is not at all important. At other places, the jacket, which, if the tube is fully loaded, has to dissipate 3.5 watts in total (2.5 W from the anode, 1 W from the cathode), becomes hardly warmer than the hand, and it can therefore be held by the fingers without discomfort, or allowed to touch the skin or the mucous membranes if tissue parts in body cavities have to be irradiated.

Jacket and cable

The position of the X-ray tube in the jacket into which it is built, is shown in *fig. 8*. The tube is connected to the high-tension source by means of a very flexible cable 1.5 metre long, with "Podur" (polyvinyl chloride) insulation and a metal braiding. The metal jacket is bonded to this braiding at one end and fixed at the other end to the anode can. Thus the anode is earthed via the braiding.

A second protective "Podur" sleeve is drawn over the metal braiding and makes a liquid-tight seal

with the metal jacket. The X-ray tube can therefore be easily disinfected, which is necessary for application in body cavities. The total diameter of this very flexible cable, built up as described, is only 8 mm.

The thin cable stands up very well to the high voltage to which it is subjected. The X-ray tube is operated by D.C. voltages up to 25 kV; it has been shown that the free cable (i.e. without X-ray tube) does not break down, even if subjected to a direct voltage of 150 kV at room temperature. When subjected to alternating current, which is always much more dangerous to cables, the test showed that the cable could stand a voltage of 60 kV peak value for thirty minutes without breaking down.

The high-tension generator

In developing the KT apparatus we have been able to avail ourselves of an earlier development in the field of television. A high-tension generator had been developed in Eindhoven for supplying the cathode-ray tube used in projection television⁵⁾, possessing exactly the properties we needed for the KT tube supply. This generator produces a well smoothed direct voltage of 25 kV at a load of $150\ \mu\text{A}$, its characteristic being such that the short-circuit current is not much greater than the normal working current, and the output capacitance is only small, so that the energy which will be discharged through the cathode-ray tube in the event of a fault (breakdown) is not excessively large. The last two properties, all important to the avoidance of overloading of the fluorescent screen, are also essential to our purpose; it is obvious that all overloading of the very small KT tube, however short, should be avoided.

We have consequently been able to copy, in broad outline, the design of the high-tension generator referred to, which has already been described in this review. In the present article, therefore, we shall give only a résumé of the principle and mention the deviations from the earlier construction which were found to be necessary.

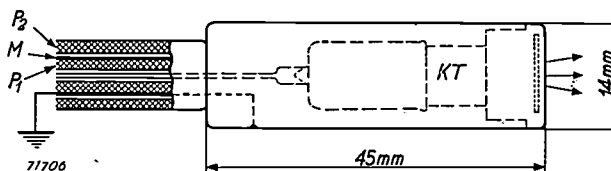


Fig. 8. Position of the X-ray tube (KT) in the jacket, and sectional view of the cable. P_1 "Podur" insulation, M metal braiding, P_2 outer "Podur" sleeve.

⁵⁾ Projection-television receiver, III: G. J. Siezen and F. Kerkhof, The 25 kV anode voltage supply unit, Philips techn. Rev. 10, 125-134, 1948.

An oscillator circuit, consisting of a coil L and its own capacitance C_p , is built into the anode circuit of a pentode. The anode current is periodically interrupted by a saw-tooth voltage applied to the grid of the pentode. At each interruption a damped oscillation occurs in the circuit with a peak value:

$$V_{\max} = i_{\max} \sqrt{L/C_p};$$

i_{\max} is the anode current at the moment of interruption. With suitable circuit constants, a voltage peak of the order of 10 kV can be obtained and the interruption can be given a rather high rate of repetition, e.g. of the order of a 1000 times a second. The intermittent high voltage in the oscillator circuit is converted by a cascade rectifier, composed of a number of valves and capacitors, into a direct voltage which is a multiple of V_{\max} , viz., in our case, 25 kV.

The rather high repetition frequency mentioned above is one of the essential items of the design. As a consequence of this high frequency the resulting direct voltage can be well smoothed with a relatively small capacitance. This amounts to about 2000 pF; the ripple in the voltage on the X-ray tube, at 200 μ A tube current, is only about 100 V.

The television generator supplies a positive voltage of 25 kV relative to earth. We needed a negative voltage for the KT tube; this can be obtained by merely reversing the valves in the cascade rectifier.

For the television tube a fixed voltage of 25 kV is required. In our case it was desired that the X-ray tube should also be operable at lower voltages. We have obtained a continuously adjustable tube voltage by making the anode voltage of the pentode variable: this involves variation of the anode current i_{\max} at the moment of interruption and, consequently, V_{\max} (see the formula given above).

In the television generator the cathode filaments of the cascade valves are fed from a winding coupled to the coil (L) in the anode circuit, and consequently supplying a fixed voltage. In our case the voltage on the coil is variable, preventing the use of this simple method; we had thus to provide for the valves a separate filament transformer, with high tension insulation. We have so arranged it, however, that this transformer is switched on and off simultaneously with the high tension. The service life of the valves is therefore not unduly shortened by useless burning. The resulting slight delay in the excitation of X-rays due to the filaments of the valves first having to be heated and the capacitors to be charged, after switching on the high tension,

was not considered objectionable in this case; it amounts to less than $\frac{1}{2}$ second. The filament transformer for the X-ray tube cathode is also switched on and off simultaneously with the above-mentioned filament transformer. Consequently no heat is generated in the tube so long as it is not under high tension and the temperature of the jacket does not rise unduly from its initial value. By means of an adjustable resistor in series with the last-mentioned filament transformer, the current in the X-ray tube can be continuously varied from 0 to 200 μ A.

Adjustments of tube current and tube voltage are not completely independent of each other; moreover, no coupling has been provided between the two adjustments to preclude the possibility of the maximum voltage (25 kV) or the maximum power (2.5 W) being exceeded. Such refinements would have made the apparatus too complicated and expensive, in view

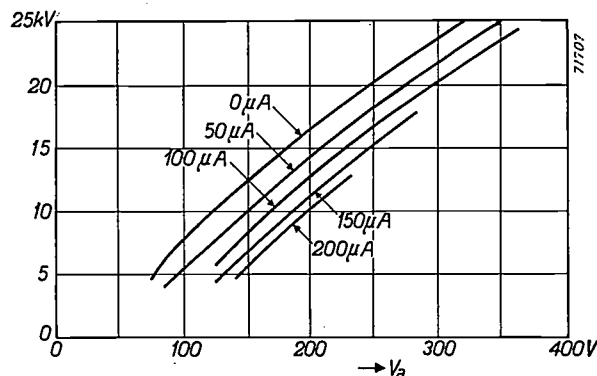


Fig. 9. Calibration curves for the voltage regulation of the KT apparatus. The high tension on the KT tube is varied by adjustment of the pentode anode voltage V_a . The relationship between the two depends also, however, on the tube current, which is separately controlled. — This calibration chart can be made identical for all KT apparatus by means of a variable resistor.

of its restricted purpose. The only consequences for the user are that he should bear in mind the limit of 2.5 W when selecting the values of the tube current and tube voltage and that he should consult a set of calibration curves when pre-setting these values (see fig. 9). If the user intends to raise the voltage whilst the apparatus is in use, he should adjust the tube current to the new, smaller value beforehand.

The high tension generator was built into a cabinet, shown in fig. 10, together with the requisite power supply apparatus, controls, switches, etc. The cable, with the attached X-ray tube, is connected to it by means of a plug, insulated for high tension, which is tightened by means of a coupling nut. This nut at the same time presses on a pin, which first connects the braiding of the cable (and consequently the anode and the jacket of the X-ray tube) with the earth of the apparatus and then closes the anode circuit of the pentode. A small cylindrical opening to hold the X-ray tube when not

in use has been provided in the lid of the cabinet (in the photograph the tube is just being withdrawn).

Fig. 11 shows the cabinet with side walls removed.

We shall merely mention here three cases for which the apparatus has proved its usefulness: (1) the irradiation of warts and haemangiomae (benign tumors of the blood vessels), more especial-



Fig. 10. The KT apparatus, ready for use. The tube current and the tube voltage are adjusted by means of the control knobs and meters. Note the connection of the "Podur" cable, with the midget KT tube at its end, to the apparatus. When not in use, this tube is housed in an opening in the apparatus.

Methods of application

Physicians who have been working with the KT apparatus for a number of months will presently report elsewhere on its use for therapeutic irradiation.

ly the so-called strawberry marks (*fig. 12*); in most cases of this type the patients are very young children, also incubator children, and the relatively small CT tube is still too large to be convenient in such cases; (2) the irradiation of the cornea of the eye,

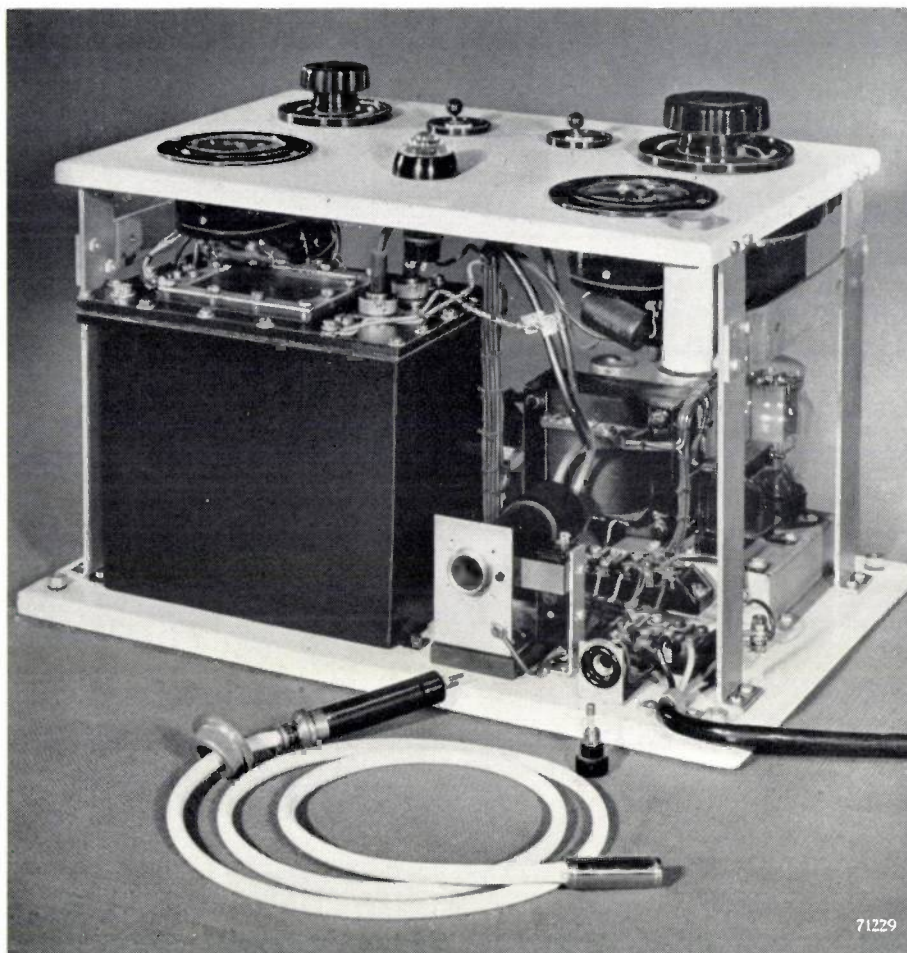


Fig. 11. The apparatus open. The black box on the left contains the high-tension coil and the cascade rectifier. On the right to the rear the power unit supplying the pentode. Shown in the foreground are the unscrewed coupling nut and the plug (withdrawn), by means of which the cable is connected. To the right of this is a safety fuse shown unscrewed.

e.g. after transplantation (*fig. 13*), and (3) the irradiation of very superficially located small tumours of the mouth and throat cavities. For each of these cases special small localizers for the limitation of the field may be useful, and aluminium filters of various thicknesses may be put in the localizers if in a special case a somewhat harder radiation is required.

The KT tube can also be usefully employed in another field, viz. the testing of materials. In support of this observation we may mention the fact (though this can hardly serve to justify its own existence!) that the KT tube renders good service in the production of — KT tubes; the rolled beryllium plate for the anode is tested for the absence of impurities, and the thin layer of gold for homogeneity, in both cases by making X-ray photographs with the KT apparatus. In general the very soft (and not unduly intense) radiation of the KT tube can be employed to good advantage for the examination of very light material or very thin layers.

Lastly, the KT tube may in our opinion be extremely useful in the biological field. An



Fig. 12. Application of the KT tube for the irradiation of a strawberry mark. (This photograph was put at our disposal by Dr. G. J. van der Plaats, radio-therapeutist at the hospital of St. Annadal, Maastricht.)

interesting example is the examination of the effect of X-rays on bacteria or other organisms present in a liquid solution. The KT tube provides a neat way of doing this; the tube is dipped in the solution which is stirred until — according to the laws of probability — each volume element of the liquid has been near the anode for a sufficiently long time to absorb the required total dose. A more obvious application of the KT apparatus is the making of

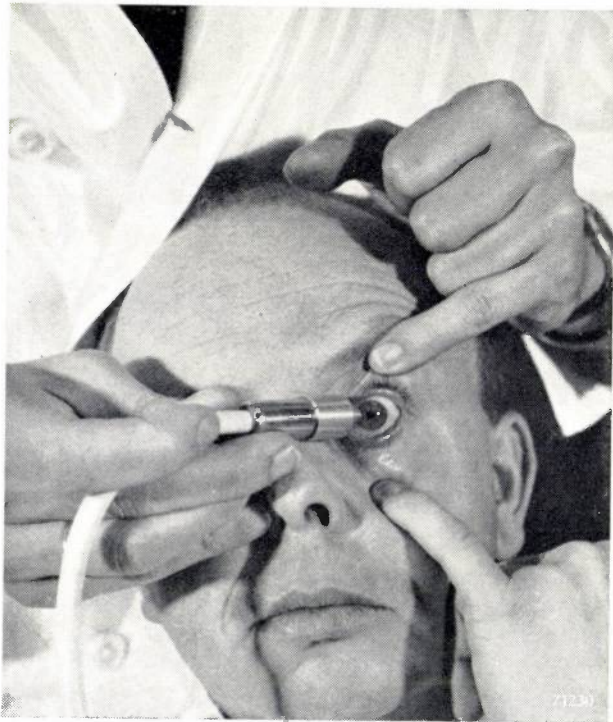


Fig. 13. Application of the KT tube for after-irradiation of a transplanted cornea. The tube is fitted, in this case, with a very small cone-shaped localiser, in order to restrict the beam of rays to the requisite small diameter. Since the depth penetration of the radiation is extremely small, special measures for protection against secondary radiation are not necessary. (The photograph was put at our disposal by Mr. P. J. L. Scholte, radio-therapist at the Municipal Hospital, the Hague.)

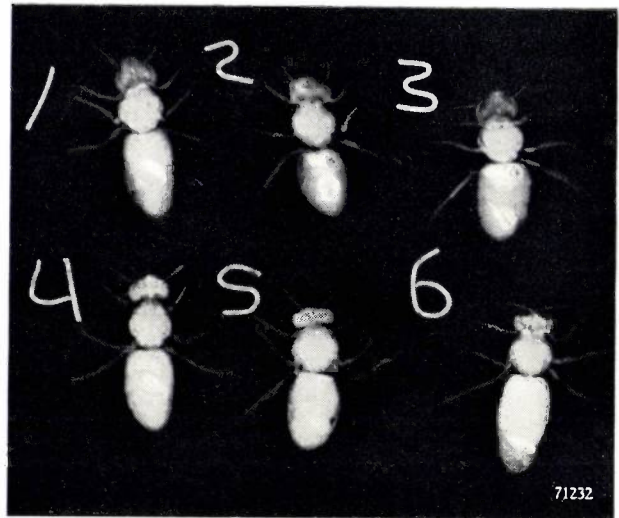


Fig. 15. X-ray photographs of honey bees after feeding them sugar water mixed with barium sulphate as contrast medium. Photographs 5, 3 and 2: half an hour after feeding with respectively 1%, 5% and 10% BaSO_4 ; photographs 1, 4 and 6 ditto an hour after feeding.

The intestines of the bee lie in its abdomen and consist chiefly of the pouch-shaped honey stomach, serving as temporary storage place for the food taken, the stomach or middle intestine, i.e. the digestive organ proper, and the rectum. The light spots seen on the photographs in the abdomen of each bee, are the honey stomach and the middle intestine. Their differences in size are due to differences in appetite; the honey stomach of bee no. 6, for example, is so well filled that its abdomen is considerably longer than that of the other bees. (The experienced beekeeper can see from the size of the abdomen if his bees collect much or little honey.) The organs of bee no. 2 are clearly shown in their normal position: the middle intestine lies in a loop. None of the six photographs shows positively the contrast medium in the rectum.

By means of photographs such as these, the absorption of the food by and its passage through the intestines of the honey bee could be studied. (The photographs were made in cooperation with Drs. A. J. de Groot, Lab. for Comparative Physiology, Utrecht, to whom we also owe this explanation.)

contact photographs of biological objects, so very thin that this could be done nearly — but still not quite — with light rays. Fig. 14 gives as an example an X-ray photograph, made with a KT apparatus,

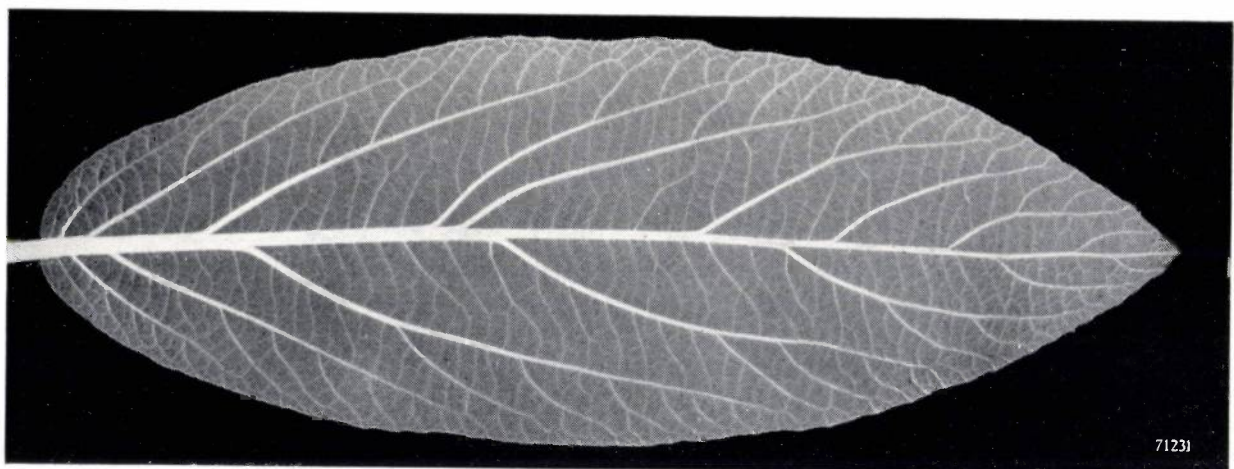


Fig. 14. X-ray photograph of a shrub-leaf, taken with the KT tube. In this case the voltage was 18 kV, the tube current 100 μA , the distance from focus to film 30 cm and the exposure time 180 sec.

of a relatively thick, barely translucent leaf of a shrub, *Fig. 15* is still more striking; it reproduces X-ray photographs of bees, after imbibing a sugar solution mixed with barium sulphate, the well-known contrast medium for radiological stomach examination; the six photographs correspond to different periods of time after feeding.

It may suffice to give these examples here.

Summary. The experimental X-ray tube described in this article (KT tube from the Dutch for smallest therapy, "kleinste therapie", tube) is 45 mm long and 14 mm in diameter, including the earthed metal jacket. The X-rays are excited in a very thin layer of gold, deposited on a beryllium plate, which at the same time serves as exit window for the rays. The tube can be loaded with a voltage of 25 kV and a power of maximum 2.5 W. The half-value thickness of the radiation is very small, viz. 0.5 mm tissue at 25 kV and 0.3 mm tissue at 10 kV, owing to the small "inherent filter" of the tube (equivalent to 1.5 mm beryllium). Very high dosage rates of some hundreds of thousands of röntgen units per minute can be obtained,

in spite of the very small power, and a rapid decrease of the dose in proportion to depth, owing to the fact that the focus can be brought within 1 mm distance of the area to be irradiated. The dose distribution across the irradiated area is very uniform; due to the method of construction, the rays that pass the window obliquely are weakened more than axial rays, but this effect is more than compensated for by the natural directional effect of the continuous X-radiation, which is a consequence of its mechanism of origin, and is effective in the KT tube owing to the extreme thinness of the gold layer.

The X-ray tube is connected to the high-tension source by means of a flexible cable, 1.5 m long and 8 mm thick, with "Podur" insulation and earthed metal braiding. The high-tension generator is, in broad outline, identical with the well-known small 25 kV generator for projection television. A low short-circuit current and a relatively small output capacitance are essential to avoid overloading of the very small tube in the event of a fault (breakdown).

In contrast to the television generator, the generator of the KT apparatus is continuously adjustable for voltages up to 25 kV and currents up to 200 μ A. The KT tube may be applied therapeutically for the removal of haemangiomas, the irradiation of the cornea after transplantation and the treatment of superficial tumors in the mouth and throat cavities. In addition, the tube can render good service in the testing of very light metals or very thin layers and for X-ray photographs of all sorts of biological objects; a series of radiographs of honey bees is, i.a., given as example.

SOME OF THE EFFECTS OF THE RELATIVE LENGTH OF DAY ON POINSETTIA AND POPULUS

by R. van der VEEN.

581.1.035.2

Systematic tests have in certain instances successfully explained some of the complex processes of plant life which depend on the amount of light they receive. It has thus become possible to say that in Poinsettia a substance which normally promotes flowering, is inactivated by an inhibitor formed in the leaves when these are exposed to long days; this substance is carried upwards and downwards, the inhibitor only downwards. This hypothesis leads to an explanation to many apparently strange instances of flowering. Similar considerations apply to the winter dormancy of poplars.

Light is an essential factor in the life of a plant or, to express it more precisely, it controls a whole system of factors. Not only are the intensity and spectral composition of the light of influence; the length of the day, by which we mean the hours of daylight, is also highly important. A discussion of these effects has already appeared in this Review ¹⁾.

Further investigation has shown that variations in the daily hours of light produce widely different effects; for example, in Poinsettia it is flowering, in the poplar the falling of the leaves that is so closely related to the relative length of the day.

In order to carry out experiments in this field it must be possible to vary the length of the "day" at will and to provide suitable reproducible conditions; we have accordingly made use of specially constructed greenhouses for this purpose ²⁾, in which plants are exposed exclusively to artificial light, and in addition are provided with an artificially controlled "climate".

The flowering response of Poinsettia

Poinsettia is decidedly a short-day plant, that is to say when it is exposed to less than 12 hours of light daily it will flower, whereas, if the duration of the illumination is more than 10 hours, one can be quite certain that it will bear no flowers. By varying the daily illumination, therefore, Poinsettia can be made flowering or vegetative at will.

When a plant which has developed during a long-day period (16 hours daylight) is given short-day treatment (9 hours light) for one week at a temperature of 26 °C and is subsequently restored to long days, it may be made to produce three similar late-

rals (fig. 1). The bud at the extremity is then a flower primordium which is, however, incapable of development. Just below this bud there will be three side shoots, which, having grown during the long days, remain vegetative. (This is the only way in which equivalent laterals can be obtained; by removing the tops, new laterals can be produced, but, as a rule, the upper ones will develop more strongly than the lower.)



Fig. 1. By growing Poinsettia under long days (16 hours), then reducing the daily light to 9 hours for one week and finally restoring the plant to the 16-hours treatment, the three equivalent laterals (1, 2 and 3) shown in the illustration are obtained.

¹⁾ R. van der Veen, Philips techn. Review **11**, 43-49, 1949.

²⁾ R. van der Veen, Philips techn. Review, **12**, 1-5, 1950 (No. 1).



Fig. 2. Poinsettia after 10 "short" days. The inflorescence comprises but few flowers and lacks the halo of red foliage normally present when the plant is in full bloom.

To ensure fully developed flowers the plant must be given at least ten short days in succession, although in that case the cluster of red leaves round the inflorescence, so characteristic of Poinsettia, is lacking; the display is scattered, with only a few flowers (*fig. 2*). After a larger number of short days more flowers are formed, but they are still rather wide apart and the inflorescence is green. After a treatment of about 30 short days there is suddenly a marked change; the inflorescence is then more concentrated and the red leaves appear (*fig. 3*). The plant thus requires some 30 short days for the induction of a complete inflorescence.

As Poinsettia shows such a pronounced response to the daily duration of the illumination, this plant was selected for the following tests.

A plant, exposed for one week to short days, produced an array of three similar laterals, one of which was removed. Of the two others one was exposed to long days (16 hours daily) and the other, by covering it for a number of hours, to short days (9 hours). As was expected, these two laterals responded independently; the short-day lateral produced flowers, whereas the other remained vegetative (*fig. 4*). The latter could nevertheless be made to

flower by removing all the leaves (*fig. 5*), doubtless owing to the influence of the leaves on the short-day branch. This test shows that — as had been demonstrated already³⁾ — the response to the relative length of the day takes place not in the growth-centre, but in the green leaves.

One single leaf left on the long-day branch was enough to prevent it from flowering. Nor were any flowers obtained on a plant, raised under long-day conditions, from which all the leaves were removed. When a plant from which the leaves have been removed is subjected to short-day treatment, the flowers again fail to appear; hence removal of the leaves is not in itself sufficient to promote flowering.

This result can be explained by assuming that under short-day conditions a flower-inducing substance is produced in the leaves, which is transported through the stem both upwards and downwards; when this substance passes a leaf which is exposed to long days it is rendered inactive by another substance formed in this leaf.

The following test confirms this assumption. A plant was taken having two laterals of the type



Fig. 3. After about 30 short days Poinsettia comes rather suddenly into full bloom; the inflorescence is compact, with a circle of red leaves round the blooms.

³⁾ See, e.g. W. S. Garner and H. A. Allard, Localisation of the response of plants to relative length of day and night *J. Agric. Res.* **31**, 871-920, 1925.



Fig. 4. Poinsettia of which the left-hand lateral has been exposed to short days (9 hours) and the right-hand one to long days (16 hours). The first is in bloom, the second purely vegetative.



Fig. 5. After removal of the leaves from the right-hand lateral in fig. 4, this also flowers under long-day conditions, owing to the influence of the left-hand lateral, which is still receiving short-day treatment.

depicted in *fig. 6*. Lateral *A* received short-day treatment, the rest long days of light. Flowers thus appeared on *A* (*fig. 6a*), but none on *C* and *D*. The leaves were then removed from *B* and *D*, but not from *C* (*fig. 6b*), for the purpose of ascertaining whether *D* would then bear flowers. Even after three months, no flowers appeared. Although en route from *A* to the growth centre of *D* not one leaf is passed which was exposed to the long-day treatment, the assumed flower-inducing substance was nevertheless inactivated on its way from *A* to *D*.

Evidently, the reason why the flowers are prevented from being formed is to be found in a substance produced in the leaves on *C* and carried downwards by way of *B*. In *B*, this substance neutralises the flower-inducing element coming from the leaves of *A*. By removing all the leaves from *C* as well, it should therefore be possible to check this neutralising effect and, actually, within one month from the time of stripping lateral *C*, buds were clearly visible not only on the extremity of *C*, but also on *D* (*fig. 6c*), notwithstanding the fact that *B*, *C* and *D* were still being exposed to long days.

It may be deduced from experiments that the flowering of Poinsettia is governed by the interaction of two substances of which the one induces flowering whilst the other renders the first inactive. It is known that auxin, the growth hormone, also tends to check flowering, and the obvious inference is that the inhibitor produced in the leaves as a

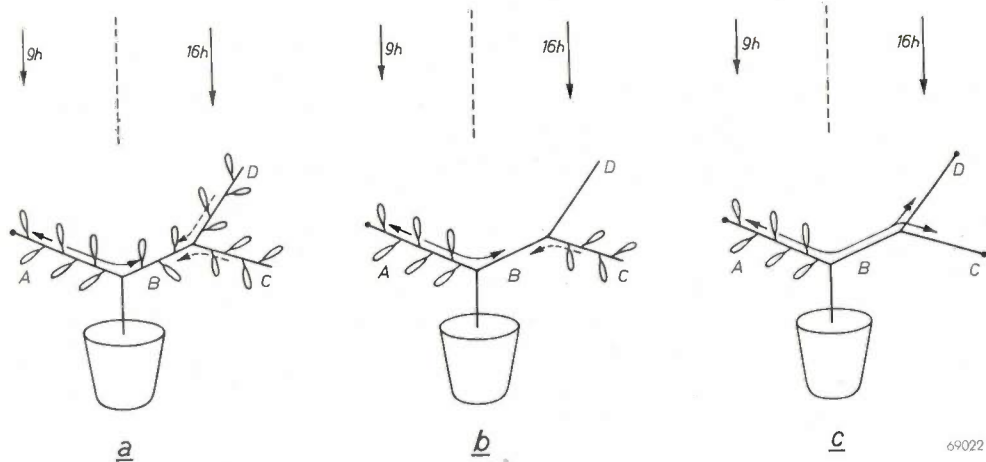


Fig. 6. Leaf-stripping tests with Poinsettia. Lateral *A* is exposed to short daily illumination (9 hours), *B*, *C* and *D* to long days (16 hours).

a) All branches vegetative; only *A* flowers.

b) *B* and *D* stripped of leaves; *D* does not flower.

c) When *C* is stripped as well, flower buds appear on *D* and *C*.

The full-line arrows indicate the movement of the flower-inducing substance formed in the leaves of the short-day lateral (*A*); the dotted arrows show the transport of the neutralising agent produced in those leaves which are exposed to long days.

result of long-day treatment is an auxin. This has not been proved, however. At the same time, it is certain that a low auxin level is not in itself sufficient to promote flowering; a positive factor is essential, namely the flower-inducing substance which is produced in the leaves during the short-day treatment.

Dormancy of Populus

In the same way that Poinsettia is a plant sensitive to the length of day from the aspect of

flowering, the poplar shows a similar response as regards its dormancy. It is only when the days are short that the leaves are shed; during the longer days growth continues. A comprehensive study of the subject has already been published⁴⁾ and we shall content ourselves with a summary of this work.

The poplar comes into growth in the spring in

⁴⁾ R. van der Veen, Influence of daylength on the dormancy of some species of the genus *Populus*, *Physiologia Plantarum* 4, 35-40, 1951 (No. 1).



Fig. 7. Two specimens of poplar (*Populus robusta*). Left: exposed to short days (12 hours) for 2 months. Right: exposed to long days (16 hours) for 2 months. The short-day plant has stopped growth. (Note the resting buds at the end of the twigs.) The growth of the long-day plant is in no way retarded.

consequence of the rise in temperature; the length of the day does not affect it. When the tree is exposed to short days, dormant buds are formed at the ends of the twigs within 6 weeks and all growth is arrested. In long days, however, the tree continues to grow (fig. 7).

If trees exposed to short-day treatment are examined four months after budding, all the leaves are found to have been shed, the tree being then quite dormant. This takes place independently of temperature; it is therefore solely due to the shortening of the days that the poplar becomes dormant (fig. 8). Plane trees respond in the same way (see fig. 9).

The first question is now in how far the branches of one and the same tree respond independently to the light conditions. To investigate this, a first-year specimen of *Populus robusta* was pruned down to two comparable branches, of which one was exposed to long days (16 hours) and the other to the short-day treatment (9 hours). In five weeks it was apparent that the short-day branch was becoming dormant, whilst the other was still growing. After three



Fig. 8. The same examples as in fig. 7, three months later under the same conditions. Although not exposed to cold, the short-day specimen (left hand) is fully dormant. The other continues to grow.

months of this treatment the last-mentioned branch still showed pronounced development, being much thicker and heavier than the other now dormant one.

This test was repeated with other specimens, with the same result. It may therefore be assumed that each branch of the poplar, too, responds independently to the lighting conditions. Fig. 9 illustrates this in the case of a plane tree.

We have made some experiments by stripping the leaves from poplars in the same way as we did with *Poinsettia*.

1) One branch exposed to short days was dormant in five to six weeks. As soon as the terminal bud was formed and dormancy was clearly evident, the leaves were removed; the dormant condition then quickly ceased and the terminal bud began to develop. In another case, where stripping of the leaves was delayed for four weeks after the onset of dormancy, the terminal bud showed no change and the tree remained dormant for another month.

Something very similar happens when a poplar, which shows the first signs of dormancy after short-day treatment, is exposed to long days; when this is done, the terminal buds soon resume their growth. On the other hand, once the tree has been dormant for a longer time, long days are ineffective.

2) We have seen that in *Poinsettia* the capacity of a short-day branch to flower can be transmitted to a long-day branch stripped of its leaves. The question is now whether from the aspect of dormancy there is a similar transmission from one branch to another in the poplar.

In order to answer this we subjected one branch of a poplar plant to the short-day treatment, at the same time stripping the leaves from the second branch. No influence of the short-day branch, however, could be observed, even after some months.

The results of these tests may be interpreted in the following manner.

Growth-promoting hormones (auxins) are formed in plants just below the growth centre and usually also in the leaves. They increase the development of the growing shoots, but they inhibit the bursting of the buds. It is well known that auxins are transported particularly in a downward direction.

The reason why the axillary buds of a normally growing branch do not burst is to be found in this auxin effect; the descending auxin inhibits the bursting of the buds. If the top of the branch is removed, however, the lack of this source of auxin considerably reduces the concentration of auxin in the region of the upper axillary buds. The inhibiting effect is thus absent and these buds will start growing, as occurs in nature when the end of a branch is cut off.

During long-day periods the production of auxin is in equilibrium with the consumption (by oxidation) of this substance, so that a certain concentration is maintained.

When the days shorten it appears that another substance is formed in the leaves which also checks

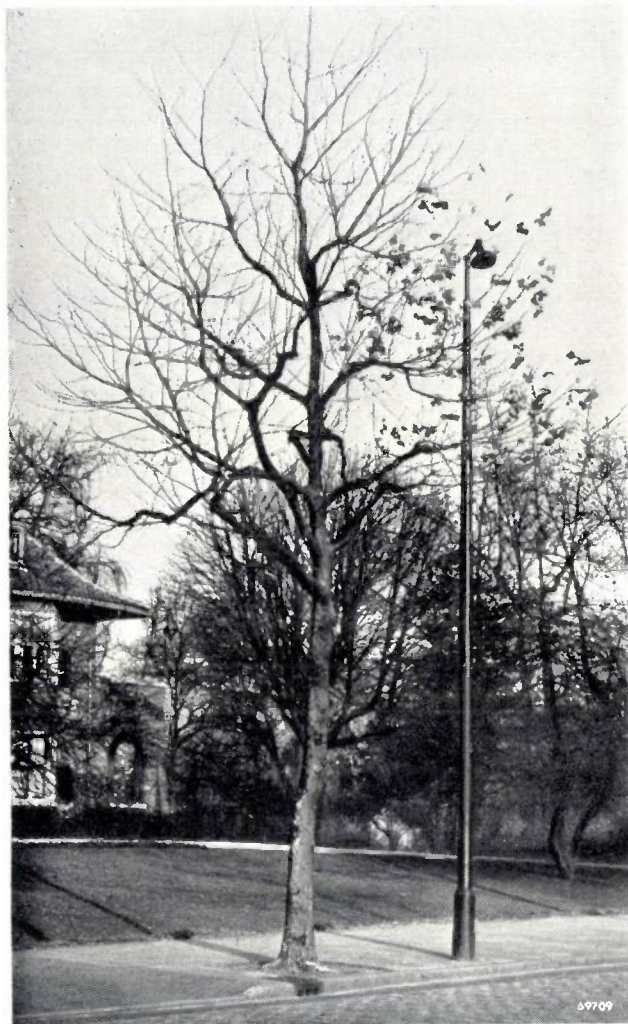


Fig. 9. Photograph of a plane tree taken at the end of November at Eindhoven. All the leaves have been shed except for that part which, owing to its proximity to the street lamp, has had the benefit of "long days".

the bursting of the buds, but in addition causes the vegetative centre to change into a dormant terminal bud. This inhibitor is, therefore, definitely transported upwards. As soon as the growth of the branch comes to an end, the flow from the auxin source just below the head ceases automatically.

This inhibitor, which induces the winter dormancy, is stored in the buds. It does not disappear as quickly as the auxin; hence the time necessary, possibly months, before the dormant branch will again become green. Everyone knows that branches of the chestnut placed in a warm room in November will not sprout, but that branches picked in February

will immediately burst into leaf.

At the onset of dormancy the surplus of inhibitor in the poplar is so small that — if this substance is no longer being produced owing to the length of the day or to stripping of the leaves — the concentration quickly drops below a certain minimum, the buds begin to grow and in the growing branch the auxins gain the supremacy.

Dormancy of longer duration results in the production of more and more of the inhibitor, so that it becomes increasingly difficult, or even impossible, to activate the buds by long days or by stripping. It is only after spontaneous leaf shedding, followed by the necessary period of rest, that the poplar is again so conditioned that the buds will open either in long or short days, the temperature being sufficiently high. During the winter dormancy the buds are particularly resistant to cold.

For the most part these considerations are not new. Murneek and others⁵⁾ have already published a comprehensive survey on the photoperiodism of plants, and Wareing⁶⁾ has given a survey of the literature of the effect of day-length on trees. The tests described above serve as good illustrations of both, having been carried out on plants which have been found to lend themselves particularly well for the purpose.

Summary. *Poinsettia* and *Populus* are plants which show a pronounced reaction to relative lengths of daylight and which are well adapted for tests in this direction. In both cases it is found that different branches of a given specimen respond quite independently to the length of the day and that "perception" of the length of day takes place in the leaves.

By subjecting one part of a *Poinsettia* to relatively short days and another to relatively long ones, and by stripping the leaves from some of the branches of the last-mentioned section, results are obtained which lend support to the following hypothesis. In those leaves which receive short daily illumination a flower-promoting substance is produced which is transported both upwards and downwards in the plant; in those leaves which receive long daily illumination an antagonistic substance or inhibitor is formed which is transported downwards only. When this latter substance encounters the flower-inducing substance, the latter is rendered inactive and those parts of the plant above the meeting point will not be influenced by the flower-inducing substance. Flowers do appear, however, at those point where the flower-inducing agent is able to penetrate without meeting the inhibitor.

In poplars, bursting of the buds is dependent on the temperature, whereas the onset of dormancy is purely a question of the duration of the daily illumination received. After five to six weeks of short days the terminal buds have been formed and the tree is dormant. The results of experiments with day-length variation and stripping of the leaves of *Populus robusta* may be interpreted as follows. In long days a certain concentration of auxin is maintained, this is transported downwards only; in short days an inhibitor is formed in the leaves which moves both upwards and downwards. This inhibitor accumulates at the buds and is inactivated very slowly.

The experiments described were carried out entirely in artificial light, under reproducible, artificial climatic conditions.

⁵⁾ A. E. Murneek, R. O. Whyte et al., Vernalization and photoperiodism, Waltham (Mass.) 1948.

⁶⁾ P. F. Wareing, Photoperiodism in woody species, *Forestry* 22, 211-220, 1947.

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

MAGNETIC SOUND RECORDING EQUIPMENT

by D. A. SNEL.

621.395.625.3

The recording of sound on magnetic material is an invention which dates back to the last century. Then it was practically forgotten, left behind by other, more perfected methods of recording; electronic amplifier techniques developed in the meantime have revived interest in it, however. In a relatively short time the magnetic system has now captured a large field of applications which were previously dependent on other methods, besides opening up new spheres in sound recording.

Introduction

In 1877 Edison surprised the world with his phonograph, an instrument for the recording and reproducing of sound, in which the vibration set up in a diaphragm by the sound waves were mechanically recorded on a steel cylinder covered with tin-foil. The cylinder was played back by means of the same apparatus. In Edison's phonograph a vertically modulated groove was produced on the cylinder, i.e. a groove the depth of which varied in accordance with the deflection of the diaphragm, in contrast with the gramophone disc introduced by the American Berliner, which is still employed today and in which the groove is modulated laterally, i.e. in the form of a curved line¹⁾. There is no need to dwell on the point that the direct, mechanical recording and reproduction of sound was later superseded by indirect methods using an electrical device (an amplifier) between the receiving element (microphone) and the cutter, as also between the needle and the reproducing diaphragm (loudspeaker).

The invention of the sound film introduced an optical method, again in conjunction with the necessary electrical appliances, in which the sound is recorded on the film photographically in the form of a track, either of constant width and variable density, or constant density and variable width; the reproduction is obtained by means of a scanning light slit and photo-electric cell.

In between the two systems of the gramophone record on the one hand and the sound film on the other, there is the Philips-Miller system which employs a sound track with variable-width modulation, cut mechanically in a celluloid tape coated with black lacquer and scanned by optical means²⁾.

Each of these methods has its own specific advantages and drawbacks, and each is employed today in its own particular sphere. In recent years, however, another system of sound recording and reproduction, based on the use of magnetic materials, has found wide application. As long ago as 1898, Valdemar Poulsen, a Dane, patented a magnetic system of recording, the sound being recorded on a steel wire wound on a drum and passed between the poles of an electromagnet. Naturally his equipment was primitive seeing that, at that time, the amplifier had not yet been invented. Poulsen's system was not developed and was accordingly shelved. About in 1935, however, when modern techniques of amplification were known, Poulsen's idea was taken up by Stille, Schüller and others in Germany; Lorenz in that country and Marconi in England placed a machine that worked with a steel tape on the market. The speed of the tape was about 1.5 m/sec, and it was possible to record frequencies up to 6000 c/s; the tape was wound on a spool about 70 cm in diameter and the playing time was a good half-hour.

¹⁾ See L. Alons, Philips techn. Rev. 13, 134, 1951 (No. 5).

²⁾ See Philips techn. Rev. 1, 107-114, 135-141, 211-214, 231-236, 1936.

As early as 1927 F. Pfleumer in Germany and Joseph O'Neill in the U.S.A. had developed paper and plastic tapes coated with a layer of magnetic material. In Germany, in the period 1940-1945, magnetic recording systems using such tapes were further developed on a large scale.

Magnetic recorders marketed nowadays are either of the "wire" or "tape" variety. The wire recorder will not be discussed here. Certainly, the equipment is usually less complicated and therefore cheaper than the tape recorder, but it has certain definite disadvantages; if the wire breaks, or if it is inexpertly handled, it will often become hopelessly tangled; the joining of two ends, either to repair a break or for the "montage" of different selections, can only be achieved by knotting, which is a much less reliable method than the splicing of two ends of tape. Several other, more fundamental, disadvantages of wire in contrast with tape are mentioned in the following paragraphs.

General construction of a tape recorder; properties of the tape

The tape is magnetized by means of a "recording head" which is in effect an annular electromagnet with small air-gap, along which the tape is passed (fig. 1). The modulation current flows in the magnet coil; this current may consist of the alternating current generated by a microphone as a result of the sound to be recorded. For the reproduction the alternating magnetization produced in the tape, the "sound-track", is scanned by a "playback" head; it contains a coil, across the ends of which an alternating voltage is produced by the passage of the tape, this voltage being amplified for passing to a headphone or loudspeaker.

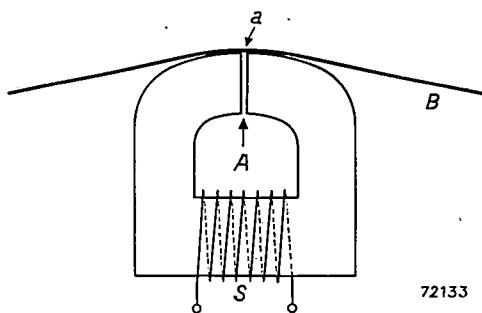
The material on which the sound is to be recorded must be capable of permanent magnetization and

must therefore exhibit a certain hysteresis and remanence; for reasons which will be explained presently, moreover, the coercive force must not be too low, say in the order of $\mu_0 H_c = 0.01 \text{ Wb/m}^2$ ³⁾. Lastly, the structure of the material must be very finely crystalline. This requirement is related to the inherent tape noise, which must always be as little as possible. Magnetization of the tape does not vary continuously from point to point, but is constant in magnitude and direction within certain small zones (Weiss zones) and changes suddenly between the boundaries of one zone and another. The noise arising from this discontinuity is so much the less according as the Weiss zones are smaller, and, since one such zone can at most extend to the confines of one whole crystal, a finely crystalline structure is obviously desirable. The simplest course is to use as basis a powdered magnetic material. As a rule, iron oxide, either red ($\gamma\text{-Fe}_2\text{O}_3$) or black (Fe_3O_4) in a grain size of less than 1μ is employed.

The powder can be incorporated in the tape in two different ways: it can be either mixed thoroughly with a thermoplastic material such as polyvinyl chloride, the tape being then extruded or rolled in the hot condition; or it can be mixed with a lacquer which is subsequently applied as a very thin layer (e.g. 15μ) to a paper or plastic tape about 35μ in thickness.

The low noise level resulting from the use of finely crystalline material is also the reason why tape is usually preferred to wire. Certainly a solid wire can be made that will exhibit very small crystals, but wire introduces another very important cause of noise which is entirely absent in tape and which will be explained in the following.

A homogeneous tape, made in accordance with the first of the two above-mentioned methods, is the most able to meet the numerous and, in some cases stringent, mechanical requirements imposed by its use in magnetic recording instruments. The chief requirements, standardized in some countries, are that the tape shall not break under loads up to 2 kg, and that the elongation one minute after removal of a load of 1 kg shall not exceed 0.2%. Further, the tape must not pull out of straight, assume a curved cross-section or shrink noticeably; it must be easily cemented and sufficiently smooth to enable it to slide over the recording head and other components without causing undue wear. The tape itself should not be subject to too much wear either; in particular it must shed as little dust as possible, must be flexible and maintain all these



72133

Fig. 1. The recording head, over which the magnetic tape passes. The modulation current flows in coil S. The tape B passes in a gentle curve over the gap A. Recording takes place at a.

³⁾ In C.G.S. units; $H_c = 100$ oersted.

characteristics against the effects of humidity and tropical temperatures.

As far as wear is concerned, it will be clear that even the least, usually unavoidable, wearing of the tape would soon fill up the air-gap in both recording and playback heads with magnetic powder and thus provide a magnetic shunt for the tape or coil. This difficulty is surmounted by filling the gap in advance with a non-magnetic material. Such material must be harder than that of which the pole pieces of the heads are made, to prevent it from wearing away and leaving an open gap. Beryllium copper is very suitable for this purpose.

Standardized dimensions for the magnetic tape are: thickness maximum 60μ , width 6.20-6.35 mm. The length is unspecified, but is usually not more than 1000 metres. The tape is magnetized either across the full width, or only for a part of it, to enable two different tracks to be recorded next to each other, or for stereophonic recordings, in which case two tracks differing only very slightly from each other are made simultaneously. In special cases, as many as six tracks have been made on one tape.

Recording with high-frequency auxiliary field; the response curve

If a magnetic particle in the tape passes the alternating magnetic field of the recording head (i.e. the gap and its immediate vicinity) so rapidly that the transit time is short compared with the periodicity of the field (i.e. of the sound to be recorded), it may be assumed that this magnetic field is constant during the time that the particle is passing through it. The particle is magnetized to a certain degree, mainly in the longitudinal direction of the tape⁴⁾ and a part of this magnetization is retained in consequence of the remanence; the remanent magnetization at any point in the tape is of course more pronounced according as the magnetic field is stronger at the moment when that point is passed.

This is not enough, however. It is necessary for the magnetization of the tape to be directly proportional to the field strength, since otherwise the well-known disturbance due to non-linear distortion (overtones and combination tones) will be manifested in the sound reproduced. Although the magnetization curve of most magnetic materials, including that of magnetic tape, is anything but linear, the required linearity can be obtained by an artifice.

⁴⁾ In principle it is also possible to magnetize the tape in the direction of the width or thickness. With the usual longitudinal magnetization a certain amount of magnetization always occurs in the direction of the thickness as well.

This consists in superimposing on the modulation current an alternating current ("auxiliary current"), the frequency of which is well beyond the audible range. The way in which the auxiliary high-frequency field thus produced accomplishes its purpose will not be discussed here; this will be explained in an article shortly to be published in this Review, dealing with the mechanism of magnetic recording in greater detail. It is sufficient to say here that, in principle, the necessary linearity can also be ensured by means of D.C. pre-magnetization; this introduces a considerable amount of background noise, however, which is absent when the high-frequency method is used.

The frequency of the auxiliary field must be so high that it changes polarity several times during the passage of a magnetic particle through it. On the other hand, it has just been pointed out that the field at the sound frequency may be regarded as constant during that time. Naturally this is no longer the case either where the highest audio frequencies are concerned. What we then record at a given point on the tape represents, as it were, an average of the field strengths to which the particle is subjected, not when passing the gap, as might be supposed, but when traversing a small zone at the edge of the gap. For a further explanation, reference must once more be made to the article announced above. One result of this formation of an average is that the high frequency components of the sound are recorded relatively weakly.

A similar effect is encountered in the playback. The alternating voltage induced in the coil of the playback head by the movement of the tape is proportional to the speed of variation of the flux in the coil. This flux is dependent on some kind of "average" of the magnetization over a length of the tape equal to the width of the air-gap. An average is thus formed in this case, too, again with detriment to the higher frequencies in the sound. The smaller the air-gap and the greater the tape speed (v), the higher the frequency at which this effect will become noticeable; the air-gap must be smaller than the shortest wavelength to be recorded and reproduced by the tape. The wavelength of a sinusoidal signal of frequency f is v/f . At the much used speed of $v = 76.2$ cm/sec (30 inch/sec) and with $f = 10\,000$ c/s, the wavelength $v/f = 76.2 \mu$. For the speed mentioned the playback head is usually given a gap of 10μ to 20μ ; the gap in the recording head may be made slightly larger, say 25μ to 40μ , this being advantageous from the point of view of more uniform magnetization throughout the thickness of the tape.

Let us now consider the response curve of the magnetic recorder as a whole, viz. the alternating output voltage as a function of the frequency of a tape recorded with constant current amplitude. Since the voltage induced in the playback coil is proportional to the speed of variation of the magnetization, it increases with frequency; in other words the response curve exhibits a natural gradient to a factor of 2 (6 dB) per octave. In the higher frequencies the increase is not so marked, and at the upper end of the range there is even a drop. In part, this is due to the above-mentioned influence of the finite dimension of the gap, but there are other causes, too. In the sound track the shortest wavelengths tend to be levelled out, seeing that closely spaced elements, magnetized with opposed polarity, have a demagnetizing effect on each other; in principle this effect should be counteracted by the iron in the playback head as soon as the relevant section of the tape approaches it, but, owing to the small intervening space that always remains between tape and head, this counterbalancing effect is not complete. (In order to suppress the effect in point as much as possible, the coercivity of the magnetic material should be high.) In the shortest of the wavelengths concerned this space between tape and head provides an opportunity for the lines of force to close up to a large extent without passing through the magnetic circuit of the playback coil. Lastly, the effect of eddy current losses in the recording and playback heads becomes more pronounced as the frequency is increased.

Whereas the response curve of the magnetic recorder accordingly assumes the form shown in *fig. 2*, a flat characteristic is our ultimate aim.

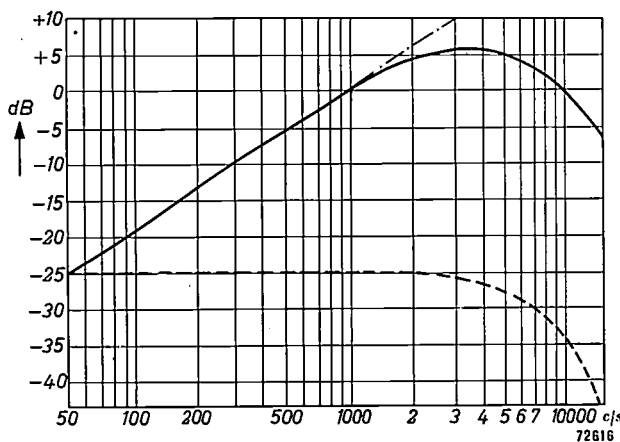


Fig. 2. Overall response curve of a magnetic recorder, showing on a relative scale the alternating voltage obtained by reproduction with constant amplitude in the recording coil, as a function of the frequency. The full line shows the response without any correction; the broken, for the greater part horizontal, line is obtained with an RC network in the playback amplifier. The higher frequencies still require boosting.

Our first course in pursuing this aim is to correct the rise in the middle frequencies. This can be achieved by means of a simple RC network in the playback amplifier, the result being the largely horizontal broken line in the figure. In the second place, the higher frequencies need to be boosted. This can be effected in both the recording and the playback amplifier. In principle it is better to boost the high notes at the recording stage, as they are then placed further beyond the noise level of the tape. This does involve the risk, however, that, in the louder passages the saturation magnetization of the tape will be reached in the higher frequencies, with consequent distortion. A compromise is therefore made, the correction being introduced partly in the recording and partly in the playback. It is a point in our favour that in orchestral music — during the recording of which the interference due to noise and distortion is the most disturbing — the strength of those components of which the frequency is high is usually less than in the middle and lower frequencies. This circumstance makes it possible to shift the greater part of the necessary correction in the higher frequencies towards the recording side.

Erasing

Recordings on magnetic tape exist by reason of the magnetized condition of the particles in the tape, and they can, therefore, be modified by a strong magnetic field, in contrast with other types of recording which, apart from the effects of physical damage, are permanent. For this reason magnetic tape recording should not be stored in places where strong magnetic fields set up by permanent magnets, transformers, motors etc. occur. The requirement that magnetic tapes should not be too sensitive to magnetic fields with which they are accidentally brought into contact is yet another reason why the coercivity of the material should not be too low (see above).

At the same time, however, the susceptibility of magnetic recording to magnetic influences is put to good use in that used tapes can be rendered magnetically neutral and thus made serviceable for fresh recordings. This erasing is effected by a separate "erasing" head, this being in the main identical with the recording and playback heads. A high-frequency magnetic field is produced in the gap of the erasing head, such that the material of which the tape is made, when passing the head, is subjected to a number of hysteresis loops which first increase and then decrease to zero on the *B-H* curve. After passing the erasing head, then, the tape is once more in the demagnetized condition.

In most recorders the sequence of operations is such that, for recording purposes, the tape first passes through the alternating field of the erasing head, thus ensuring a perfectly "clean" basis for the recording.

In the case of wire recorders it is much more difficult to erase a recording completely. This is due to the fact that the wire is drawn through slots in the recording and erasing heads, and that side of the circular cross-section of the wire which faces the open end of the slot is only slightly exposed to the recording and erasing fields. Since wire is always more or less subject to twisting, after erasing, certain parts of the wire will, generally speaking, retain traces of the previous recording. After repeated recordings and erasures all these residues tend to produce a background babble, which constitutes extra noise, not occurring in tapes.

heads and induction from external magnetic fields at mains frequency (hum). The gaps in the recording and playback heads must be precisely at right angles to the tape; any obliqueness would again affect the higher frequencies⁵⁾.

The tape driving mechanism

For a satisfactory quality of reproduction the speed of the tape passing the recording and playback heads must be highly constant. The mechanical problems to be solved in order to meet this and certain other requirements connected with the tape drive are worthy of being discussed⁶⁾. In this we shall base our considerations mainly on the design of Philips' tape recorders.

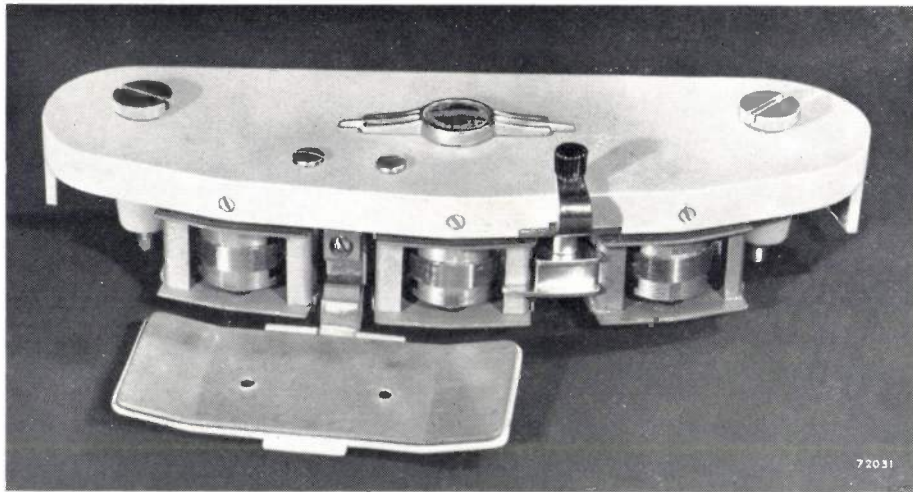


Fig. 3. The magnetic heads of a Philips tape recorder, arranged as a single unit (see fig. 6); from right to left the erasing, recording and playback heads. In front of the two last-mentioned heads a mu-metal screen is provided, this being placed over the tape to eliminate interference due to cross-talk and possible alternating fields in the vicinity.

The frequency of the erasing field is normally 35 or 40 kc/s and the auxiliary recording field generally somewhat higher, e.g. 100 kc/s. The larger magnetic recorders will usually incorporate two separate oscillators to provide these frequencies, but other equipment may have only one oscillator generating the two frequencies required, as harmonics of a fundamental frequency.

In the larger equipment the erasing, recording and playback heads are mounted in a common, detachable, holder. Fig. 3 shows the unit as fitted to one of Philips' tape recorders, this assembly being fitted with plugs for direct connection to the amplifiers. The playback and recording heads are housed in compartments of high permeability material such as mu-metal, in order to limit cross-talk between the

In principle, uniform motion of the tape is obtained by means of an accurately machined spindle which is driven by a synchronous motor and against which the tape is pressed by a rubber roller (fig. 4).

⁵⁾ Actually, it is only necessary for the gaps in the recording and playback heads to be parallel to each other. In order to ensure proper reproduction of a recorded tape on any suitable instrument, perpendicular gaps should of course be used. — It may be added here that, in the case of "amplitude" recordings, as in the Philips-Miller system as well as in one of the photographic methods, oblique positioning of the gaps results in non-linear distortion of the sound; see J. F. Schouten, Philips techn. Rev. 6, 110-119, 1941. In our own case (intensity recording) there is only linear distortion (variation in the response curve, as mentioned in the text).

⁶⁾ See also J. J. C. Hardenberg, The transport of sound film in apparatus for recording and reproduction, Philips techn. Rev. 5, 74-81, 1940, in which article similar problems concerning the drive of the Philips-Miller system are discussed.

This driving spindle must of course be made of non-magnetic material and should be as little as possible subject to wear.

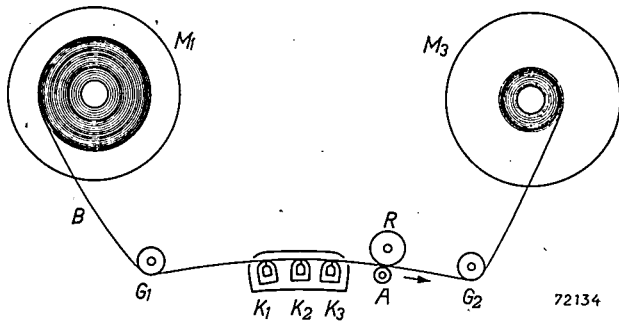


Fig. 4. Principle of the tape drive. *A* = transport spindle; *B* = magnetic tape; *R* = rubber roller holding the tape against the transport spindle. Erasing, recording and playback heads *K*₁, *K*₂ and *K*₃ are assembled in a common unit. *M*₃ = winding spool; *M*₁ = supply spool. *G*₁ and *G*₂ = guide rollers.

Eccentricity is reduced to a point below the tolerance limit of about 4μ by grinding the spindle in its own ball-bearings. The rubber roller must be highly uniform in composition and the pressure such that the tape will just be prevented from slipping when a certain tensile force (about 800 gr) is applied.

Synchronous motors ensure the roller to be rotated at a perfectly constant speed as an average, but there is always a certain slight oscillation superimposed on this rotation, owing partly to the finite number of poles in the motor and partly to a natural tendency to "hunt"; in accordance with the load on the motor, the rotor lags in phase with respect to the stator field, and this phase angle, generally speaking, does not adjust itself aperiodically, but in an oscillatory manner. This tendency is suppressed as much as possible in the design of the motor, a discussion of which would take us too far afield, however. Any residual effects, including irregularities in running due to the poles, are further reduced by placing a flywheel on the transport spindle and by transferring the motion from the motor by means of a flexible coupling ("mechanical filter").

When recording or reproduction is to be commenced, the transport spindle must quickly assume its correct working speed, viz. within a matter of seconds. In order to ensure this, and to permit the motor to come easily into synchronism with the mains, the motor must have a certain starting torque; this will have to be so much the greater according as the inertia of the flywheel is increased. It is therefore not advisable to use a heavier flywheel than is necessary to keep fluctuations in speed just within the tolerance, for which limits have been standardized internationally; the motor can then be suit-

ably proportioned to ensure that the correct speed is obtained within the desired space of time (in Philips magnetic recorders about 2 seconds).

The Comité Consultatif International de Radiodiffusion have laid down the following limits for fluctuations in the speed of magnetic recorders intended for broadcasting purposes: max. 1 per thousand at 30 inch/sec tape speed, max. 1.5 per thousand at 15 inch/sec, max. 5 per thousand at 7.5 inch/sec.

Special equipment has been developed by means of which, when reproducing a recorded frequency of 3000 c/s, a direct reading as a percentage can be obtained of variations in speeds, if desired split into fluctuations of frequency < 20 c/s (wow) and > 20 c/s (flutter).

The synchronous motor driving the transport spindle must not fall out of synchronism when the mains voltage fluctuates within the extremes of the prevailing values. In some equipment, moreover, it must be suitable for two tape speeds, in which case the dimensions of the flywheel should be such that speed fluctuations will not be too great at the lower speed, the motor being such that, with this (relatively heavy) flywheel, it will develop the proper speed quickly enough for the higher of the two tape speeds as well⁷⁾.

During recording and playback, not only the transport spindle has to be driven, but also the spool on which the tape is wound up. Simultaneously the supply spool from which the tape is being unwound must be braked in order to ensure that the tape remains taut; this is essential in order to keep the space between the tape and the heads, as mentioned above, sufficiently small (fig. 4). For rewinding purposes the supply spool has to be driven, and, to prevent looping of the tape, the other spool is then braked. As a rule, higher speeds are required for rewinding than for recording or playing; it is also convenient to be able to run the tape through faster in the forward direction when collecting particular passages of recordings for montage purposes.

In the larger magnetic recorders all these conditions are met by providing separate motors for the two spools. For recording and playing both are energized with a fraction of the mains voltage; the working spool then slowly winds up the tape and the emptying spool receives a slight torque in the oppo-

⁷⁾ In recorders that work with a slow tape speed the motor can be run with the tape disengaged from the transport roller; the tape can then be started up immediately at any given moment by springing it suddenly against the roller. If the tape speed is on the fast side, however, this involves too much risk of snapping the tape; for that matter, even in slow-speed instruments suitable means have to be provided for absorbing such jerks.

site direction, so that the tape is held taut. For rapid running or rewinding the motor is fully energized to give a tape speed some 10 or 20 times faster than normal.

The tensile forces exerted on the tape by the motors and the inevitable slight fluctuations in these forces must obviously affect the uniformity of the tape movement past the recording and playback heads. The outcome is a complex system of forces, the effect of which can best be studied with the aid of an equivalent electrical circuit. This is shown in fig. 5, a brief explanation being given in the caption.

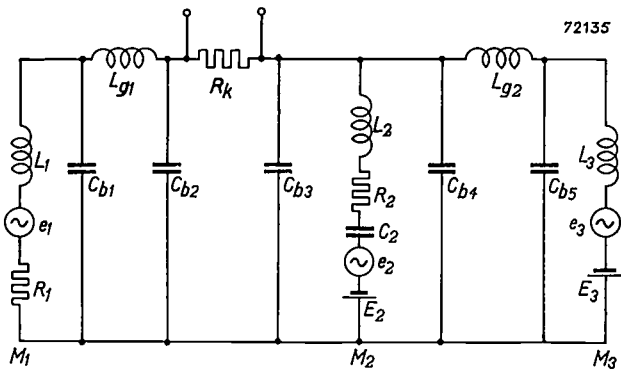


Fig. 5. Equivalent electrical circuit of the mechanical tape drive, by means of which the influence of the various components on the speed of the tape can be investigated. A rotating mass is represented by a choke (L), a constant force by an e.m.f. (E), fluctuations in the force by an alternating voltage (e), a friction or a constant braking force by a resistor (R) and a spring force by a capacitor (C). In particular, M_2 = synchronous motor (represented by E_2 and e_2); L_2 = flywheel with rotor and transport spindle; C_2 and R_2 = spring force and internal friction of the flexible coupling; M_3 = motor for winding spool (with E_3 and e_3); L_3 mass of rotor and winding spool; M_1 motor of supply spool (with braking forces R_1 and e_1); L_1 = mass of rotor and supply spool; R_k = friction of tape against the heads; L_{g1} and L_{g2} mass of guide rollers; C_{b1} , C_{b2} , C_{b3} , C_{b4} and C_{b5} = resilient sections of the tape.

The tape speed at the head is represented by current flowing in R_k ; in the ideal case this would be a direct current without any ripple. To this end the ripple voltage from e_2 must be fully smoothed by L_2 , C_2 , R_2 and C_{b3} ; e_1 is largely smoothed by L_1 , the remainder is smoothed by the "filter" C_{b1} , L_{g1} , C_{b2} ; similar conditions apply to e_3 .

The forces applied by the spooling motors (regarded as constant for the purposes of the diagram and represented by the e.m.f. E_3 and resistance R_1) will in general depend on the diameter of the roll of tape wound, or left, on the respective spools. In principle, this could result in a variation in the tape speed from the beginning to the end of the tape. It is possible to design the motors so that the forces will be independent of the diameter; experience has shown, however, that a certain amount of variation in the force, to the extent of say 1:4, is quite permissible. For such a limitation of the variation it suffices to make the spool cores large enough.

It should be noted here that the type of drive just described, in which constant speed is ensured mainly by the transport spindle, can be used only in tape recorders, not with wire. In wire-recording, variations in speed are therefore usually appreciably greater, and a gradual change in the speed from the beginning to the end of the recording is by no means so easily prevented. This is particularly troublesome when different parts of a recording are extracted and assembled; during the playback a sudden change in the pitch of the sound may then be audible.

Rapid spooling and rewinding of the tape introduces another problem of its own, namely the possibility of stopping the tape suddenly when a wanted passage is found. Both the spools must then be capable of stopping quickly and simultaneously if the tape is not to be allowed to run slack, or to be stretched possibly to the breaking point. Numerous devices have been tried out. In Philips' recorders this problem is solved by means of carefully balanced brake-bands on the spools, actuated by springs which are released by Bowden cables when a push-button is depressed to stop the motors. This system is very reliable, besides being superior to, and more flexible than the electromagnetic braking system often employed, which necessitates special measures to avoid audible interference due to induction effect in the tape. To prevent unnecessary wear, a slide is pushed up when the tape is to be run through quickly, thus holding the latter clear of the erasing and recording heads (this can be seen between the heads in fig. 3); the tape, however, maintains light contact with the playback head, so that the sound can be more or less followed and a wanted passage rapidly located.

Some current models of tape recorder

Of the different models of magnetic recorder made by Philips let us now consider two with a view to explaining some of the special features, viz. a large model for professional use in broadcasting studios etc. (type No. 10028/03) and a small portable "semi-professional" set, which is at the same time suitable for amateur use (type No. EL 3540).

The first of these is depicted in fig. 6. There are two tape speeds, viz. 76.2 and 38.1 cm/sec which, with a standard tape length of 970 m corresponds to a playing time of 21 and 42 minutes respectively. The response curve is flat to within plus or minus 2 dB between 30 and 15 000 c/s and the noise level (background, hum etc.) is 54 dB below signal level on full output, i.e. the level at which distortion is 2%. Variations in the tape speed, both transient and throughout the length of the tape, are within 0.1% at the speed of 76.2 cm/sec and within 0.15% at 38.1 cm/sec.



Fig. 6. Large magnetic recorder for broadcasting and film studios, etc., type No. 10028 03. On the top panel will be seen the tape spools, the head assembly and the control knobs. The oscillators and voltage supply units are housed inside the cabinet.

The entire equipment is contained in a steel cabinet. The driving mechanism is mounted under the top panel, on the upper side of which will be seen the two spools, the recording, playback and erasing unit and the press-button controls (see also *fig. 7*). The cabinet is fitted with two drawers which slide out to the rear and contain the recording amplifier, the playback amplifier with correcting network, the two high-frequency oscillators (100 kc/s for recording and 40 kc/s for erasing) and a common voltage supply unit. The drawers are equipped with built-in plug pins to facilitate their removal and replacement for servicing purposes.

To ensure that the whole unit shall be as versatile as possible, the high-frequency and speech currents are both variable; by means of two switches (lower panel, *fig. 6*), each of these can be adjusted to any one of 11 stages. A control is also provided for

boosting the high tones to a variable extent during recording, so that optimum playback quality can be obtained whatever the characteristics of the magnetic tape used.

The upper front panel contains a meter from which can be read the values of the recording and playback currents, as well as the high frequency and erasing currents. The recording and playback amplifiers are designed to receive and deliver the signal at "line" level, i.e. a level that is suitable for transmission by telephone line to a broadcasting station. A jack in parallel with the meter serves for connection of headphones for monitoring the recording or playback.

Owing to the very fine tolerances employed in the workmanship, the sensitivity of the recording and playback heads — as well as their other characteristics — vary very little. Defective heads can, therefore, be replaced without any readjustment, to give the same output level within about 1 dB.

Fig. 8 depicts the other model under review. This is of much simpler design and has only two heads, one for erasing and one for recording and playback; a single oscillator (pre-magnetization and erasing are both effected at 45 kc/s); only one motor, which



Fig. 7. Close-up of the top panel of the magnetic recorder depicted in *fig. 6*. The components of the driving mechanism sketched in *fig. 4* are easily recognized. At the front of this panel, on the left, will be seen the push-buttons for the control. The first of these (left to right) is for the rewind; rotation of the knob on the extreme left controls the speed, both forward and reverse. The second button starts the recorder at normal speed forward; this is the "playback" setting, seeing that the playback head is permanently connected to the amplifier (all amplifiers are of course previously switched on). The third push-button brings in the erasing head and recording head. To avoid the possibility of erasing part of a recording by depressing the button accidentally, this button is interlocked in such a way that it functions only when depressed together with the second button (which merely starts the machine). The fourth button returns all the others to neutral and stops the instrument by braking the motors.

by mechanical means fulfils all three of the functions of normal drive, fast forwards and fast rewind; finally, recording and reproduction are both carried out by means of the same amplifier: one output pentode is used for both functions by means of a switching manipulation, however, on the understanding that for the playback the pentode of the oscillator, which would otherwise be inoperative, is connected in push-pull with the first-mentioned valve in order to further improve output quality.

Amplifier distortion at 1 W output is below 2%; at 6 W output below 5%. This set is equipped with an electronic indicator for checking modulation depth of modulation, and the noise level in the reproduced sound can, if desired, be further reduced at the expense of the higher tones, by means of a tone control. With a view to the use of this instrument by amateurs, its operation (loading and unloading etc.) has been kept as simple as possible.

A special feature which we should mention in

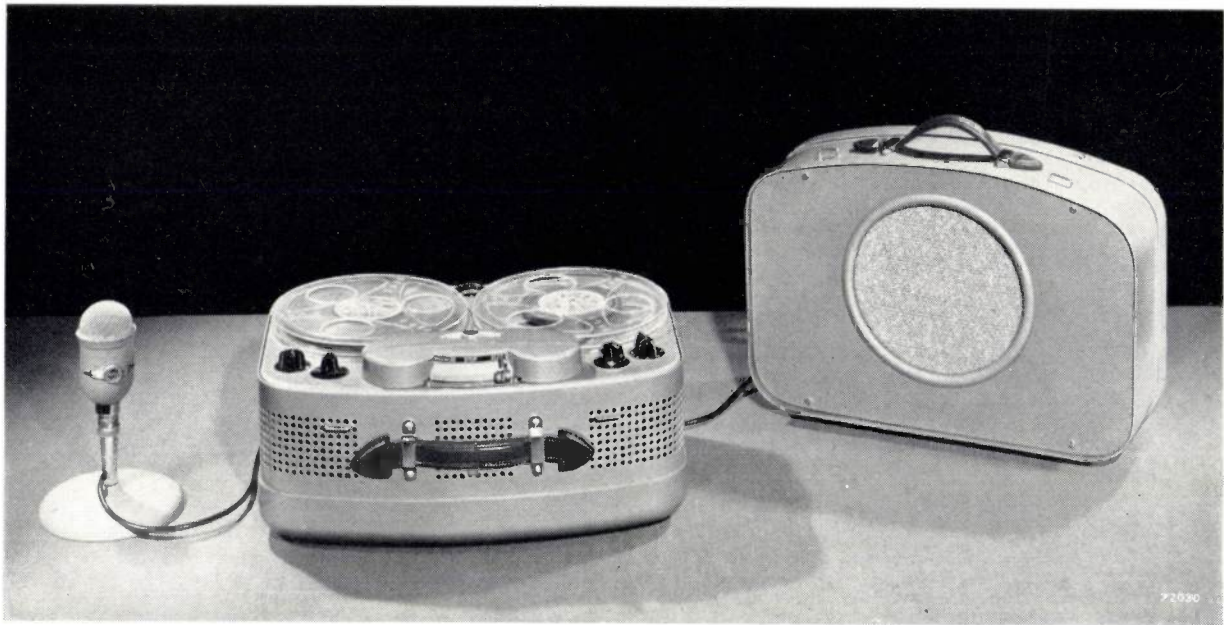


Fig. 8. Small portable magnetic tape recorder type No. EL 3540 for professional or amateur use. At the left the microphone; at the right the loudspeaker. The recorder itself is housed in one small carrying case.

By thus making the most of the limited number of components, an instrument is obtained that ensures quite good reproduction and, besides, offers all the usual operational features. As the amplifier includes an output stage, the instrument is suitable for use with much lower input voltages and at higher output voltages than in the large model; hence, if desired, the speech current can be recorded direct from a microphone and the tape can be played back direct through a loudspeaker. This tape recorder works with a speed of 19 cm/sec and the spools hold about 515 m of tape, giving an uninterrupted playing time of 45 minutes. The response curve is flat to within plus or minus 3 dB between 60 and 6000 c/s and the noise level is 50 dB below signal level on full drive. Transient variations in tape speed do not exceed 0.15% and, although the over-all variation is rather more than in the larger unit, it is still on the low side, viz. about 0.5%.

closing this article is the device for demagnetizing the recording/playback head. Each time the set is switched over from recording to playing, a current impulse is produced across the coil in the head, which results in a certain, albeit weak, remanent magnetization. The same thing happens if too strong a signal overloads the head during recording, or if a steel object, say a screw-driver, is brought too close to the head. As remanent magnetization of the head introduces distortion and noise in the recording, it is essential to demagnetize the head. In the semi-professional tape recorder under review this is effected automatically each time the change-over switch is operated; when the button is depressed the recording/playback head is temporarily disconnected and placed in contact with a charged capacitor. This capacitor discharges a damped oscillation through the head, thus fully demagnetizing it, regardless of the extent of the spurious magnetization.

In professional tape recorders with separate recording and playback heads, an impulse of this kind is not produced, so that there is less need to guard against other sources of magnetization, occasional demagnetization in the usual way by means of a decaying alternating field being sufficient.

Summary. A description of the principle of modern magnetic-tape sound recording. The composition of the tape is discussed from the point of view of the magnetic and mechanical

requirements to which the tape has to conform; this is followed by a general review of the manner in which the recording is made, using an auxiliary high frequency field, the overall response curve and its correction and the erasing of the recorded sound track. Details of the tape drive are given, in particular the means of ensuring very constant speed for both recording and playing, as also for rapid running and rewinding for "montage", checking and so on. In conclusion two of the tape recorders made by Philips are reviewed, viz. a large unit for use in broadcasting studios etc., and a small portable instrument suitable for both professional purposes and amateur use. The more important features of these two apparatus are mentioned, together with some special details of the design.

A STABILISED EXTRA-HIGH TENSION RECTIFIER FOR 5000 V, 50 mA

by P. PERILHOU † *) and J. CAYZAC *).

621.314.67:621.316.722.1

A direct voltage that at the maximum variations in load varies only to the extent of roughly 0,01%, that is subject to fluctuations equal to not more than 1/3000th part of the mains voltage variations and that is variable within wide limits; these are the outstanding features of a stabilised rectifier designed specially for laboratory work in the field of decimetric and centimetric waves.

There are many instances in which a rectifier is found preferable to a battery (with charging equipment), since it demands no regular maintenance. Moreover, especially when high voltages are required, a rectifier is usually very much less voluminous and cumbersome and is therefore more easily transportable. Nevertheless, if a particularly constant voltage is needed, rectifiers have this disadvantage that the direct voltage is subject to relative variations equal to those inherent in the mains supplies to which they are connected, unless special precautions are taken to prevent this. Added to this is the fact that the internal resistance of the conventional type of rectifier is usually higher than that of a battery of comparable output. The output voltage of a rectifier is thus generally more dependent on the load than is the case with batteries.

Stabilised rectifiers which are capable of meeting this difficulty were introduced considerable time ago; these include a control circuit that will maintain a sufficiently constant output voltage on a fluctuating mains supply and/or load to an even

greater extent than a battery working on varying loads. More than 10 years ago this Review included an article on the subject of circuits suitable for this purpose¹⁾, the principle of which is illustrated with reference to fig. 1.

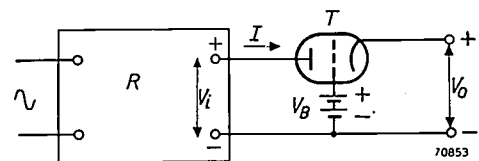


Fig. 1. Circuit for producing a stabilised direct voltage V_0 . R = rectifier with smoothing filter, delivering non-stabilised voltage V_i . T = control valve. I = direct current. V_B ($\approx V_0$) = constant reference voltage supplied by a battery.

In this figure R represents a rectifier with smoothing circuit. The output voltage V_i is dependent on the mains voltage and the load current I . The control circuit comprises a triode T , known as the control valve, which functions as a variable resistor in series with the load, in the positive D.C. line. A battery of which the voltage V_B is roughly equal to the desired output voltage V_0 , is connected between the grid of the valve and the negative line, the polarity being such that the difference $V_B - V_0$ serves as grid bias.

*) Laboratoires d'Electronique et de Physique appliquees, Paris. M. Pierre Périllhou, head of the Philips Laboratoires Hyperfréquences at Paris, died on 11th April, 1950, as the result of an accident. It was on his initiative that the work that forms the basis of the present article was undertaken, and the greater part of this work was carried out under his guidance. The article was written after the death of M. Périllhou.

¹⁾ H. J. Lindenhovius and H. Rinia, A direct current supply apparatus with stabilised voltage, Philips techn. Rev. 6, 54-61, 1941.

In professional tape recorders with separate recording and playback heads, an impulse of this kind is not produced, so that there is less need to guard against other sources of magnetization, occasional demagnetization in the usual way by means of a decaying alternating field being sufficient.

Summary. A description of the principle of modern magnetic-tape sound recording. The composition of the tape is discussed from the point of view of the magnetic and mechanical

requirements to which the tape has to conform; this is followed by a general review of the manner in which the recording is made, using an auxiliary high frequency field, the overall response curve and its correction and the erasing of the recorded sound track. Details of the tape drive are given, in particular the means of ensuring very constant speed for both recording and playing, as also for rapid running and rewinding for "montage", checking and so on. In conclusion two of the tape recorders made by Philips are reviewed, viz. a large unit for use in broadcasting studios etc., and a small portable instrument suitable for both professional purposes and amateur use. The more important features of these two apparatus are mentioned, together with some special details of the design.

A STABILISED EXTRA-HIGH TENSION RECTIFIER FOR 5000 V, 50 mA

by P. PERILHOU † *) and J. CAYZAC *).

621.314.67:621.316.722.1

A direct voltage that at the maximum variations in load varies only to the extent of roughly 0,01%, that is subject to fluctuations equal to not more than 1/3000th part of the mains voltage variations and that is variable within wide limits; these are the outstanding features of a stabilised rectifier designed specially for laboratory work in the field of decimetric and centimetric waves.

There are many instances in which a rectifier is found preferable to a battery (with charging equipment), since it demands no regular maintenance. Moreover, especially when high voltages are required, a rectifier is usually very much less voluminous and cumbersome and is therefore more easily transportable. Nevertheless, if a particularly constant voltage is needed, rectifiers have this disadvantage that the direct voltage is subject to relative variations equal to those inherent in the mains supplies to which they are connected, unless special precautions are taken to prevent this. Added to this is the fact that the internal resistance of the conventional type of rectifier is usually higher than that of a battery of comparable output. The output voltage of a rectifier is thus generally more dependent on the load than is the case with batteries.

Stabilised rectifiers which are capable of meeting this difficulty were introduced considerable time ago; these include a control circuit that will maintain a sufficiently constant output voltage on a fluctuating mains supply and/or load to an even

greater extent than a battery working on varying loads. More than 10 years ago this Review included an article on the subject of circuits suitable for this purpose ¹⁾, the principle of which is illustrated with reference to fig. 1.

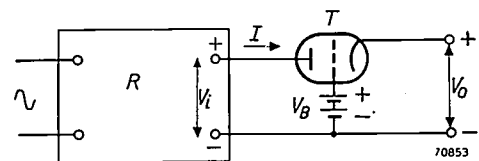


Fig. 1. Circuit for producing a stabilised direct voltage V_0 . R = rectifier with smoothing filter, delivering non-stabilised voltage V_i . T = control valve. I = direct current. V_B ($\approx V_0$) = constant reference voltage supplied by a battery.

In this figure R represents a rectifier with smoothing circuit. The output voltage V_i is dependent on the mains voltage and the load current I . The control circuit comprises a triode T , known as the control valve, which functions as a variable resistor in series with the load, in the positive D.C. line. A battery of which the voltage V_B is roughly equal to the desired output voltage V_0 , is connected between the grid of the valve and the negative line, the polarity being such that the difference $V_B - V_0$ serves as grid bias.

*) Laboratoires d'Electronique et de Physique appliquees, Paris. M. Pierre Périllhou, head of the Philips Laboratoires Hyperfréquences at Paris, died on 11th April, 1950, as the result of an accident. It was on his initiative that the work that forms the basis of the present article was undertaken, and the greater part of this work was carried out under his guidance. The article was written after the death of M. Périllhou.

¹⁾ H. J. Lindenhovius and H. Rinia, A direct current supply apparatus with stabilised voltage, Philips techn. Rev. 6, 54-61, 1941.

Briefly, the working is as follows. The voltage V_0 finds a level that is slightly higher than V_B , such that the grid is just sufficiently negative with respect to the cathode to allow the required current I to flow. Increases in V_0 raise the negative potential of the grid and, conversely, decreases in V_0 reduce it, so that any variation in V_0 , whatever the cause of it may be, is at once counteracted. A simple calculation will show that the attenuation factor α , that is, the ratio of the relative variation in V_i to that in V_0 , is nearly enough equal to the amplification factor μ of the control valve (see the article quoted in note 1)). This amplification factor may be, say, 20.

Very similar considerations apply to the internal resistance, i.e. the quantity that indicates the extent to which V_0 varies in consequence of changes in the strength of the current. As shown in the article referred to in note 1), the internal resistance at the output terminals is roughly μ times lower than that of a rectifier without control circuit.

The higher the amplification factor μ , the less the dependence of the output voltage on the input voltage and on the current. The amplification factor of triodes is, however, limited to values which are not particularly high. In order to make the stabilisation as accurate as possible, it is therefore customary to employ a direct voltage amplifier (A , fig. 2) to amplify the difference between the battery voltage V_B and the output voltage V_0 by a factor n . The output voltage is then used as grid bias for the control valve; instead of μ , we now have $n\mu$, the attenuation factor is n times higher and the internal resistance n times lower than without the amplifier. It is a simple matter to obtain a value of n of 100, which represents a considerable improvement compared with the original conditions.

The position is less satisfactory, however, if the output voltage V_0 is so high that it would be considered a disadvantage to have to use a sufficiently large number of batteries to ensure that the reference voltage V_B would actually approximate to V_0 . It is then preferable to work with a lower value of V_B and to compare this lower value with a fraction (p) of V_0 (fig. 3). This means sacrificing some of the accuracy of the compensation, viz. to the extent of $1/p$ and, in order to restore the original degree of accuracy, the amplification must be increased by $1/p$.

If the voltage pV_0 is taken from a variable potentiometer in the manner shown in fig. 3, the output voltage V_0 becomes continuously variable, for V_0 will in every instance adjust itself so that pV_0 is approximately equal to V_B . V_0 is

thus varied by adjusting p . (In principle the same purpose is served by varying V_B ; for continuous control, however, it would then be necessary to connect a potentiometer across at least a part

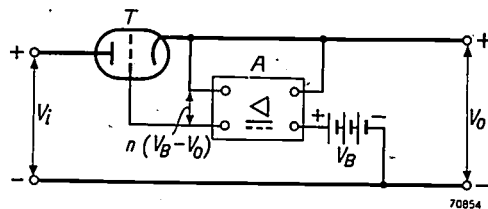


Fig. 2. Circuit as in fig. 1, but with direct voltage amplifier A to amplify the small difference in potential $V_B - V_0$ by a factor n .

of the battery, and this part would be under constant discharge; hence this is not a very suitable method, especially when dry batteries are used. It may be noticed that in figs 1, 2 and 3 the battery supplies no current.)

So much, then, for the principle on which the voltage stabilisation is based. Further details and variations, as well as a description of a stabilised rectifier for 300 V, 100 mA, will be found in the article referred to in note 1).

Special circumstances may necessitate special solutions. For example, if an unusually high direct voltage and only a small amount of current are required — which means a very high load resistance — stray capacitances have a detrimental effect. A solution for a problem of this kind has already been described in an earlier issue of this Review²⁾ (acceleration voltage, for an electron microscope, of 50 to 100 kV, at only a fraction of a milliampere). In the following a particular instance will be described, viz. that of a direct voltage

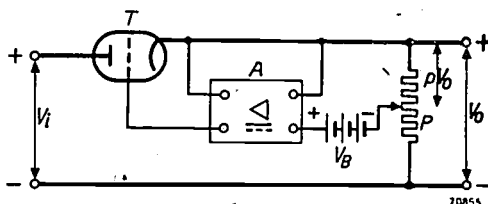


Fig. 3. As fig. 2, but with V_B approximately equal to the voltage pV_0 tapped from V_0 by means of a potentiometer.

source required for laboratory work in the field of decimetric and centimetric waves. The velocity modulation valves employed to generate these frequencies are such that the frequency is dependent on the voltage supplied to the valve. As the frequency must be very constant indeed, variations

²⁾ A. C. van Dorsten, Philips techn. Rev. 10, 135-140, 1948.

in the supply voltage must be kept to an extreme minimum. The special feature of this problem lies mainly in the amplification circuit, which, owing to the high direct voltage, differs somewhat from the conventional arrangement.

Main requirements to be fulfilled by the rectifier

The following are the conditions imposed on a stabilised direct voltage source of the kind envisaged:

- 1) A direct voltage continuously variable from 900 to 5000 V.
- 2) Direct current: maximum 50 mA.
- 3) Relative fluctuations in the direct voltage to be at most 1/1000th part of the relative mains voltage fluctuations, which may amount to 10% of the nominal value.

triode TB 2.5/300³); the rectifier proper consists of a transformer Tr_1 , two rectifier valves and a smoothing circuit. Reference to the amplifier will be made later.

As an output voltage of 5000 V is required, and as about 1000 V is lost in the control valve, the voltage V_i at the output of the smoothing filter should be at least 6000 V. Apart from the required direct current of maximum 50 mA the rectifier should be capable of supplying an extra 15 mA for feeding the amplifier, the potentiometer P (and other resistors, the purposes of which will be mentioned presently), as well as a voltmeter shunted across the output terminals.

The required flexibility between 900 and 5000 V is obtained by means of a control in large and small

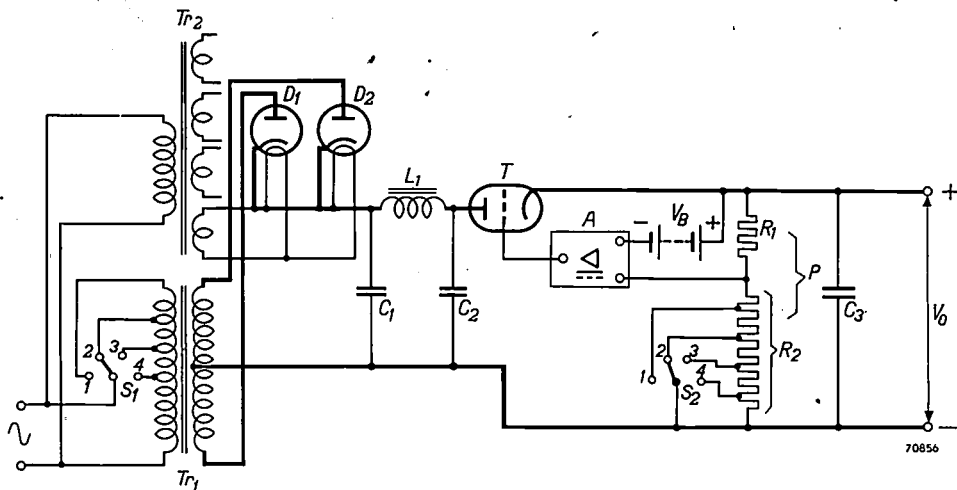


Fig. 4. Circuit of stabilised rectifier for 5000 V, 50 mA, Tr_1 = H.T. transformer. S_1 = tapping switch for adjustment of the secondary voltage of Tr_1 in four stages. Tr_2 = heater transformer. D_1, D_2 = diodes. C_1, L_1, C_2 = smoothing filter. P = potentiometer comprising fixed resistor R_1 and tapped resistor R_2 with switch S_2 (ganged to S_1). C_3 = output capacitance. V_B, A and T as in the previous circuit.

- 4) At 5000 V, variations of 50 mA in the current must not result in voltage variations greater than 0.5 V (0.01%); in other words, the internal resistance (for D.C.) should be less than 10 Ω .
- 5) The internal impedance to frequencies of some hundreds of kilocycles must not exceed 20 Ω (at 5000 V).
- 6) On 5000 V the combined noise and ripple voltages must not exceed 10 mV.

The methods by which these conditions are met are described below.

Description of a stabilised rectifier to deliver 5000 V, 50 mA

Fig. 4 shows the main details of the circuit. The control circuit is similar to that depicted in fig. 3, the control valve being the transmitting

steps, combined with continuous adjustment. Switch S_2 (fig. 4) provides the rough adjustment by varying the resistance of R_2 in stages. R_2 and a fixed resistor R_1 together constitute the potentiometer P ; when S_2 is operated the fraction p tapped from the voltage V_0 and, with it, the voltage V_0 itself is varied, seeing that pV_0 is at all times approximately equal to the constant reference voltage V_B (approx. 300 V).

In order that the control valve will not have to absorb a potential difference of several kilovolts when a low value of V_0 is required, say 1000 V, the secondary voltage from the transformer Tr_1 is modified simultaneously with the rotation of switch S_2 , by means of switch S_1 which is ganged

³) Described in Philips techn. Rev. 10, 273-281, 1948, in particular pp. 278-280.

to it. For the four positions of these switches, V_0 assumes values between the following limits:

- S_1 and S_2 in position 1: V_0 between 900 and 2100 V,
- S_1 and S_2 in position 2: V_0 between 1900 and 3100 V,
- S_1 and S_2 in position 3: V_0 between 2900 and 4100 V,
- S_1 and S_2 in position 4: V_0 between 3900 and 5100 V.

As will be shown presently, the arrangement is rather more complicated, seeing that when the switches are operated, not one but three resistances have to be varied.

Control in small stages (200 V) as well as the continuous control (over a range of at least 200 V) are effected by smaller tappings on the resistor R_2 , and by varying a portion of R_2 , respectively. These two methods of control are not shown in fig. 4.

The amplifier

The question now arises: what must be the amplification n of the amplifier A to ensure that relative mains voltage fluctuations appear in the direct voltage V_0 1000 times reduced?

We have seen that $n\mu$ is equal to the desired attenuation factor a if the battery voltage is taken to be equal to V_0 , but that, with a lower battery voltage, the gain must be $1/p = V_0/V_B$ times greater. To this must be added a factor V_i/V_0 , to make allowance for the fact that a certain change in V_i will represent V_i/V_0 times as great a percentage of V_0 as of V_i ; hence:

$$n\mu = a \frac{1}{p} \frac{V_i}{V_0} = a \frac{V_i}{V_B}$$

With $a = 1000$, $V_i = 5000 + 1000 = 6000$ V and

$V_B = 300$ V, it follows that:

$$n\mu = 20\,000.$$

As the amplification factor μ of the triode TB 2.5/300 is 20, n must be 1000.

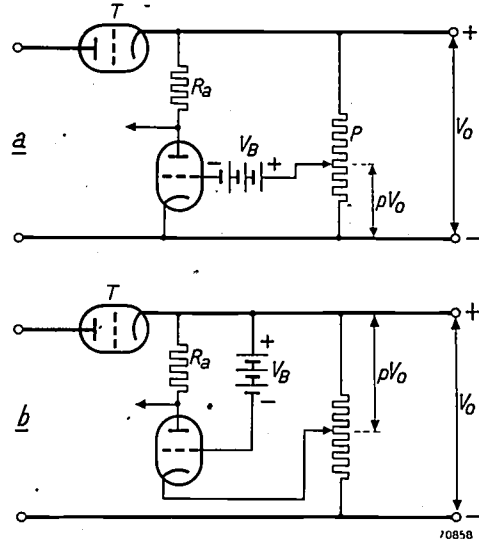


Fig. 6. a) Circuit of the first valve of the amplifier A (fig. 4). This circuit is practicable only when V_0 does not exceed a couple of hundred volts. b) In this case the first amplifier valve is on the positive side; the lower end of the potentiometer P now serves as a biasing resistor and produces negative feedback, which may be strong enough to reduce the gain in this stage to about unity.

This cannot be achieved with a single stage of amplification; two stages are necessary and these yield a gain of 3000; hence the condition that a must be at least 1000 has been amply fulfilled.

The first stage (fig. 5) comprises a triode (section II of a double triode ECC 40), and the second a pentode (EF 42), the amplification factors of these valves being 25 and 120 respectively. The other half of the ECC 40 serves as an input stage with no gain but ensuring freedom from certain difficulties.

If V_B is to be constant, the battery must be connected in such a way that it will supply no current, i.e. in series with the grid of a valve. This valve could be the first in the amplifier — connected in the manner shown in fig. 6a — if the voltage V_0 were not more than two or three hundred volts. If V_0 is much higher, however, it becomes difficult to provide a direct voltage coupling between the first valve (the electrodes of which are roughly at the potential of the negative line) and the second amplifier valve, which has to deliver the grid bias for the control valve and is therefore roughly at the same potential as the positive line.

This can be overcome by adopting the method illustrated in fig. 6b, in which the first valve is also placed on the positive side. The negative end of the potentiometer P now functions as a biasing resistor and also introduces a certain amount of negative feedback, which reduces the gain. When V_0 is high, this resistance is so high that the gain of the first

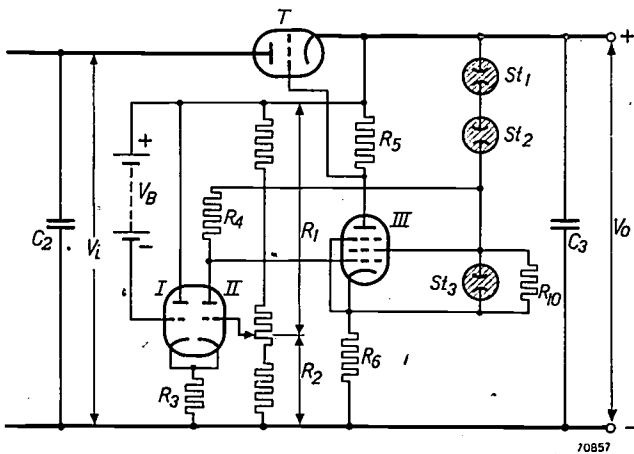


Fig. 5. Circuit diagram showing some of the details of the direct voltage amplifier A in fig. 4. Section I of the double triode ECC 40 works as a cathode follower and section II as first amplifier valve. The second valve, III, is a pentode EF 42. R_3 and R_6 are biasing resistors. R_4 and R_5 are anode load resistors. St_1 , St_2 , St_3 = stabiliser tubes type 85A1, supplying anode voltage for triode II and screen voltage for pentode III. R_{10} is referred to later (fig. 13). Other references as in fig. 4.

valve is reduced approximately to 1. Two further stages are thus necessary for the required gain of 1000.

In the suggested circuit (fig. 5) there is an input stage without gain, followed by two stages of amplification, with this difference as compared with fig. 6b that the two sections of the ECC 40 constitute a Miller compensator. The feature of this circuit is that variations in the cathode emission (e.g. due to fluctuations in the heater current caused by mains voltage fluctuations) are largely compensated⁴⁾. For this reason special stabilisation of the heater current is unnecessary.

Section I of the ECC 40 works as a cathode follower: the potential of the cathode follows that of the grid, i.e. the potential V_0 , by way of the battery. When V_0 increases, for example, the potential of the cathodes of this valve rises by the same amount, whereas the grid potential of section II increases only slightly. The anode of II and the control grid of the EF 42 thus acquire a higher potential, the current flowing in the load resistor R_5 rises, the grid of the control valve becomes more negative and thus counteracts the increase in V_0 . In this way, then, variations in V_0 have the effect of adjusting the bias of the control valve in the right direction.

The extreme values which this bias must be able to assume are -7 and -170 V and the anode current of the EF 42 must be capable of variation between corresponding limits.

The amplifier is fed from the direct voltage terminals, and three stabiliser tubes type 85A1⁵⁾ in series ensure that the voltages supplied to the amplifier valves are constant.

To simplify assembly, the circuit is so arranged that no high potential differences with respect to the positive line occur at any point in the amplifier itself; the cathodes of the ECC 40 are at about -300 V and the cathode of the EF 42 carries about -255 V with respect to positive. The remainder of V_0 is taken up by the resistors R_3 and R_6 (fig. 5).

Adjustment of the direct voltage

It has already been stated that the direct voltage can be controlled in large and small steps as well as continuously, the main control in the range from 900 to 5100 V being in four overlapping steps of 1200 V. This is achieved by means of switches S_1 and S_2 (fig. 4), which provide four values of the secondary voltage from the transformer Tr_1 , and four potentiometer ratios p . Simultaneously with p (i.e. resistor R_2), the resistors R_3 and R_6 must be suitably varied, for which purpose switches S_1 and

S_2 are ganged to S_3 and S_4 (fig. 7), R_3 and R_6 being so varied that the correct amount of current flows in each of these resistors (4.45 mA, which also applies to R_2). A fifth switch, S_5 , also ganged to S_1 and S_2 , sets the voltmeter across the output terminals for a measuring range of 4000 V in positions 1 and 2, and for a range of 6000 V in positions 3 and 4.

A second set of ganged switches, S_6 , S_7 and S_8 , provides adjustment of the voltage in six steps of 200 V.

Further, R_3 , R_2 and R_6 can be varied by means of ganged, continuously variable resistors, giving a gradual control over a range of 200 V. In this way any required direct voltage between 900 and 5100 V is at once obtainable.

Prevention of instability

As will be seen from fig. 2 onwards, the amplifier and control valve form a closed circuit. Under certain conditions this circuit is liable to oscillate. For example the following undesirable situation might arise when the H.T. transformer is switched on. Before the switch is closed, no current flows through R_5 (fig. 5), the bias of the control valve is therefore zero and the resistance of this valve (anode to cathode) is low. Consequently, when the circuit is made, the direct voltage V_0 rises rapidly, particularly if the load is inductive. The current flowing through R_5 then also rises, although, when there is some delay in the amplifier, less quickly. It may then happen that, when V_0 reaches the required value, the grid bias (across R_5) will not have acquired its corresponding value, with the result that V_0 rises beyond the appropriate level, to the point where the bias becomes so highly negative that the control valve forms a high resistance. V_0 then drops, but because of the delay in the amplifier the potential across R_5 arrives too late to be effective and V_0 is reduced too much, and so on. Instead of attaining equilibrium, V_0 thus fluctuates about the required value.

It would take us too far afield to enter into all the details of the precautions necessary to prevent such instability, and only brief mention of these will be made.

Everything possible has been done to reduce delaying influences, i.e. phase displacements in the amplifier; hence the absence of coupling capacitors and grid leaks. In this way the frequency at which instability might set in is shifted towards the higher values (beyond 15 000 c/s). A choke (L_2 , fig. 7) having a high impedance at that frequency, is connected in series with the control valve.

⁴⁾ S. E. Miller, Sensitive D.C. amplifier with A.C. operation, *Electronics* 14, 27-31 and 105-109, Nov. 1941.

⁵⁾ T. Jurriaanse, A voltage stabilizing tube for very constant voltage, *Philips techn. Rev.* 8, 272-277, 1946.

(In order to prevent this choke from increasing the internal impedance too much, it is shunted by a 10 000 Ω resistor). A capacitor C_3 across the output terminals serves to check abrupt variations in V_0 (and also ensures a low internal impedance at high frequencies, viz. always less than 20 Ω).

The potentiometer P consists of wire-wound resistors; these are not non-inductive, and the

at the output to the grid and cathode of triode *II* without phase displacement, and with a polarity such that the fluctuations are largely compensated. It is therefore not objectionable that a fairly high ripple is present on the output side of the smoothing filter or, in other words, this filter (C_1 - L_1 - C_2 , figs 4 and 7) need not conform to any special requirements. With the values of $C_1 = C_2 = 0.5 \mu\text{F}$ and $L_1 = 40 \text{ H}$

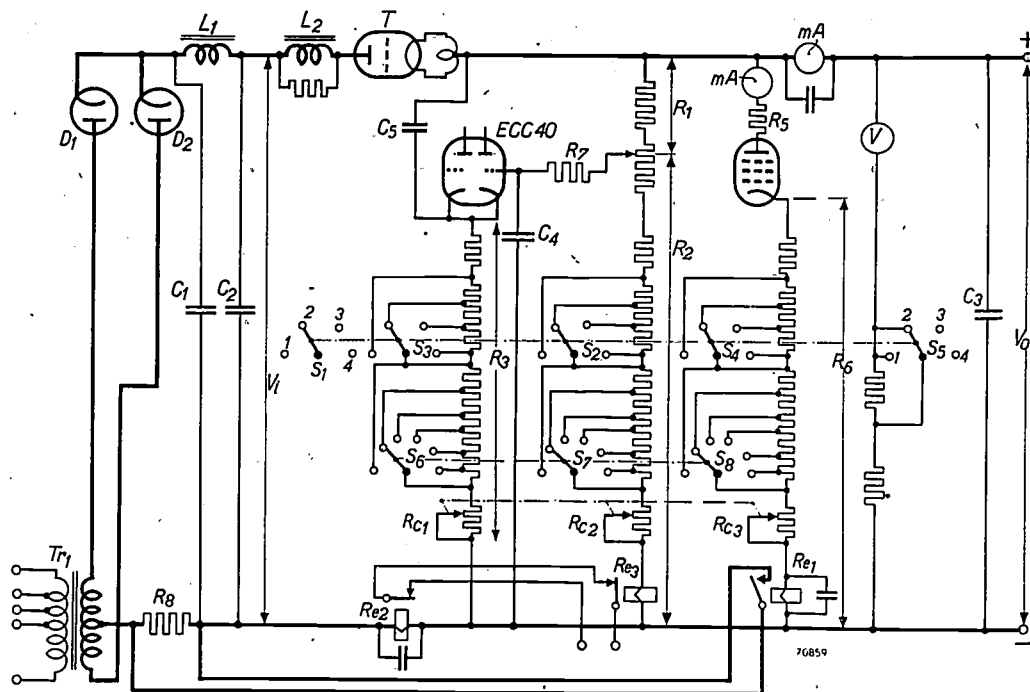


Fig. 7. Ganged switches S_1, S_2 (cf. fig. 4), S_3, S_4 provide adjustment of the voltage V_0 between 900 and 5100 V in stages of 1200 V. Each such stage is divided into further steps of 200 V by means of ganged switches S_5 - S_7 - S_8 . Within each of these smaller steps V_0 can be varied continuously by the ganged variable resistors R_{c1}, R_{c2}, R_{c3} . Switch S_5 (ganged to S_1 - S_2 - S_3 - S_4) sets the voltmeter V to a test range of 4000 V in positions 1 and 2, and 6000 V in positions 3 and 4.

To prevent instability: L_2 (8 H), output capacitor C_3 (0.2 μF), decoupling circuit R_7 - C_4 . Protective measures. To avoid transient current surges in the rectifier valves when the generator is switched on: resistor R_8 (subsequently shorted by relay Re_1). Overload protection: relay Re_2 . Excess voltage protection: relay Re_3 . When Re_2 or Re_3 opens, the H.T. transformer circuit is broken (see fig. 8).

Other references as in previous diagrams.

potentiometer thus tends to produce a phase displacement which favours instability. The decoupling circuit R_7 - C_4 (fig. 7), however, prevents any alternating voltage component in the voltage tapped from P from reaching the grid of triode *II*.

These precautions are sufficient to suppress any tendency towards instability.

The smoothing

The control circuit ensures that not only random, but also regular fluctuations in the input voltage — the ripple — appear in the output in an attenuated form. The capacitors C_4 and C_5 (fig. 7) pass possible alternating voltage components

as employed by us, the ripple across C_2 , with $V_0 = 5000 \text{ V}$ and at full load, is about 30 V. The corresponding ripple (including noise) at the output is 1/3000 of this value (10 mV).

Protective devices

Rectifier valves with oxide-coated cathodes suffer damage if anode current is allowed to flow before the cathode has been properly heated; a safety circuit of the kind usually employed in rectifiers containing such valves has therefore been adopted in the present case.

This circuit is depicted in fig. 8, which shows the use of a bimetallic switch to permit the closing

of the H.T. transformer circuit only when the heater transformer has been switched on for a certain time (45 sec). Details are given in the caption of the figure.

A second device for saving the cathodes consists in preceding the smoothing filter by a resistor R_8 (fig. 7), which prevents excessive current from flowing in the rectifier valves immediately the H.T. transformer is switched on, at which moment the smoothing capacitors are not yet charged. Some time after the circuit has been made, that is, when the voltage V_0 has risen to the point where it

the control valve TB 2.5/300, as far as possible from other components, in a corner of the cabinet (fig. 10), and to check thermal radiation by means of polished nickel screen plates.

The reference voltage is supplied by three series-connected dry batteries of 103 V nominal rating and of a current type. These batteries last for years and have a very low temperature coefficient⁶⁾; they are housed in a metal box (B , fig. 10) lined with insulating material. The negative lead is carried out by means of a cable with high-quality insulation (polythene).

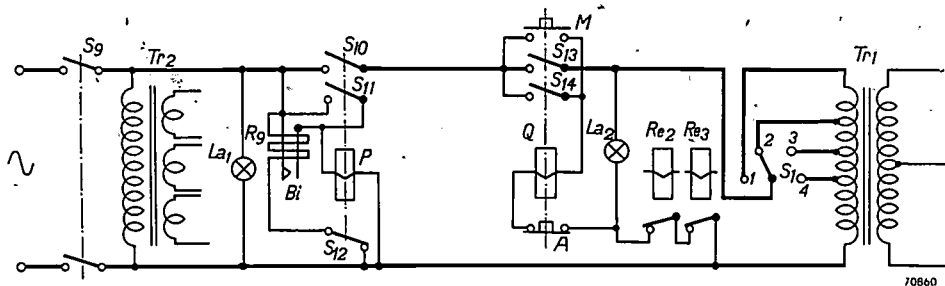


Fig. 8. Circuit for automatically preventing current from flowing in the rectifier valves before the oxide-coated cathodes have reached the required temperature. S_9 = switch between the heater transformer Tr_2 and the mains. Current passes through the resistor R_9 wound on the bimetallic strip Bi which deflects and closes contact after 45 sec. Magnet coil P of the electro-magnetic switch is thus energised, i.e. the main contact S_{10} and auxiliary contact S_{11} close and S_{12} opens. S_{12} breaks the current path through R_9 , the bimetallic strip cools and breaks its contact (in the meantime shunted by S_{11} , so that P remains energised) and after a time is again ready for a cycle of operations. Only when S_{10} is closed will pressure on the push button M allow the coil Q of a second electromagnet switch to be energised; the main contact S_{13} closes the H.T. transformer circuit. When M is released, Q remains energised through the auxiliary contact S_{14} .

The H.T. voltage can be switched off without interrupting the heater current, either by hand (push button A) or automatically (relay Re_2 or Re_3 ; see fig. 7).
 La_1 , La_2 pilot lamps.

can operate a relay Re_1 , the resistor R_8 is short-circuited.

If the delivered direct current becomes too high, another relay, Re_2 (fig. 7), opens and thereby breaks the H.T. transformer circuit (fig. 8). Furthermore, should the control circuit become defective, resulting in excessive values of V_0 , relay Re_3 (fig. 7), through the coil of which flows a current that is proportional to V_0 , similarly breaks the H.T. transformer circuit.

Constructional features

The complete unit is depicted in fig. 9 (exterior) and fig. 10 (interior), and further details are given in the captions. To these may be added a few remarks relating to heat and insulation.

Holes in the bottom and sides, and an air-gap under the top cover provide adequate ventilation of the cabinet. None the less it was considered desirable to mount the main source of heat, viz.

"Plexiglas" was used to mount the amplifier the control valve holder, etc., in the cabinet.

Resistors, the values of which are critical, viz. R_1 , R_2 , R_3 and R_6 (fig. 7), are made from constantan wire, wound on quartz bushes in order to avoid leakage currents. The wire is wound in such a way that voltages induced in them by possible stray fields counteract each other.

Fig. 11 shows the method of earthing either the positive or the negative side by means of a special plug, so arranged that the outer conductor of the cable is automatically earthed and the inner core connected to the non-earthed pole (all insulations are calculated for both contingencies).

⁶⁾ With some sacrifice in accuracy it is possible to avoid the use of batteries by deriving the reference voltage from a number of voltage reference tubes type 85A1 or 85A2 in series (see article referred to in note⁵⁾). The voltage variation in these tubes during life is not more than 0.5% and their temperature coefficient is about -3 mV/ $^{\circ}$ C. In both respects the tubes closely approximate to the performance of the batteries of the type employed.



Fig. 9

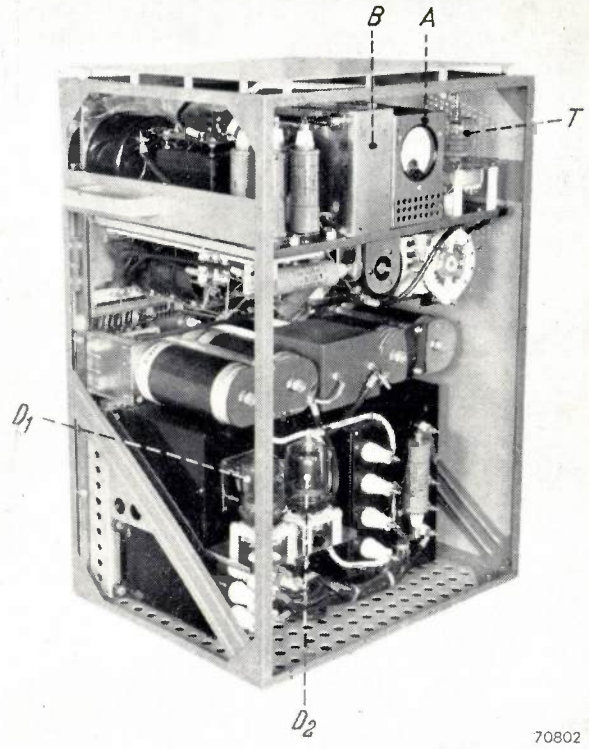


Fig. 10

Fig. 9. Front view of the rectifier. From left to right and from top to bottom: voltmeter for the direct voltage, milliammeter for the D.C., controls for the direct voltage (rough, vernier, continuous), pilot lamp, push buttons operating S_0 and S_{13} - S_{11} (fig. 8), second pilot lamp, fuses on A.C. side (one is reversible for 110 and 220 V), H.T. cable, mains lead. Consumption on full load 700 W. Dimensions: base 445 mm \times 350 mm; height 790 mm. Weight 80 kg.

Fig. 10. Interior of the rectifier (side and back panels removed). A = amplifier (with milliammeter in anode circuit of the EF 42). B = battery box, T = control valve TB 2.5/300. D_1, D_2 = diodes.

The transient state immediately after the generator is switched on

When the H.T. transformer is switched on, the amplifier operates for a time under abnormal conditions. The output voltage V_0 rises gradually from zero and, until it attains the value at which the stabiliser tubes ignite, these tubes pass no current and no anode current flows in valve II

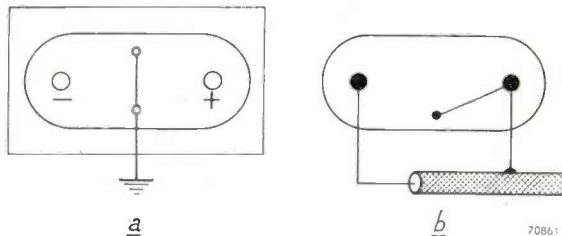


Fig. 11. a) The direct voltage amplifier sockets. b) Corresponding plug on the H.T. cable. In one position of the plug the negative pole and outer conductor of the cable are earthed, with the positive pole to inner core; with the plug reversed, the positive pole and outer conductor are earthed and the negative pole is connected to the core.

(fig. 5) and neither — as the screen-grid voltage is zero — in valve III.

This abnormal condition entails some risk that a state of equilibrium may be reached in which the value of V_0 is not by any means the desired one. The manner in which such a spurious state of equilibrium may arise is demonstrated in fig. 12, in which curve I was plotted using a variable direct voltage U_g from a separate source ⁷⁾ as grid bias for the control valve, instead of the potential across R_5 (fig. 5). A fixed load resistance R_0 was connected across the output terminals. The input voltage V_i was maintained at a constant value by adjusting the primary alternating voltage. Curve I, fig. 12, represents U_g as a function of V_0 . For a value of 900 V for V_0 the curve shows that, with the given values of R_0 and V_i , $U_g = -59$ V (point P). If U_g be now replaced by the potential across R_5 , the question arises as to the manner in which the voltage U_{R_5} will vary immediately

⁷⁾ U is assumed to represent the potentials with respect to the positive line, V those with respect to the negative line.

after the generator is switched on, i.e. when V_0 rises from zero. If this takes place in accordance with curve 2, which does not intersect 1 before the point P , equilibrium will occur at the desired value

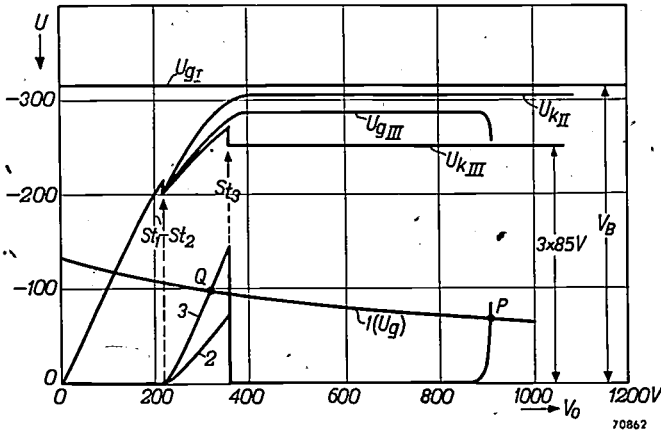


Fig. 12. Curve 1: grid bias U_g required for the control valve at $V_i = 3000$ V and load resistor $R_0 = 18\ 000\ \Omega$ to yield the values of V_0 plotted as abscissae. Curves 2 and 3: voltage U_{R5} across resistor R_5 (fig. 5) as a function of V_0 . Curve 2 intersects 1 only at point P , and V_0 reaches the required value of 900 V. Curve 3 intersects 1 at a previous point Q , and V_0 finds a state of equilibrium at too low a value.

The other curves represent the potentials U_{kII} and U_{kIII} , U_{gI} ($= -V_B$) and U_{gIII} (see fig. 13) as functions of V_0 . The arrows indicate the values of V_0 at which the tubes St_1 , St_2 , St_3 ignite.

of V_0 (900 V). But if U_{R5} should develop in accordance with curve 3, which cuts curve 1 not only at P , but at a previous point Q , equilibrium will be established at Q , that is to say, V_0 will assume a value that is much too low. It is therefore important to know U_{R5} as a function of V_0 from $V_0 = 0$ onwards.

That this function will assume the shape of curve 2 or 3 in fig. 12 is at once apparent from fig. 5. So long as V_0 does not reach the value at which the series-connected stabiliser tubes St_1 and St_2 ignite — which takes place at about 210 V — the pentode, being without screen voltage, passes no current, and U_{R5} is zero. From the moment that St_1 and St_2 ignite, the pentode passes both screen and anode current and the latter also flows in valve II. The anode current of the pentode depends on the magnitude of the voltage between the grid and the cathode, that is, the difference between the potential U_{gIII} of the control grid and the potential U_{kIII} of the cathode. These potentials, both with respect to the positive line, are also shown as functions of V_0 in fig. 12. When V_0 attains the value at which the third stabiliser tube ignites, the voltage across this tube drops suddenly from the ignition voltage (100-125 V) to the working voltage (85 V), and the difference between U_{gIII} and U_{kIII}

is increased by 15 to 40 V, that is, to the point where the pentode is cut off and U_{R5} again drops to zero.

What we now have to guard against is that the curve $U_{R5} = f(V_0)$ does not prematurely intersect curve $U_g = f(V_0)$ (fig. 12), which means that the voltage U_{R5} for igniting St_3 , and hence also the anode current of the pentode, must be kept sufficiently low, i.e. that the difference between U_{gIII} and U_{kIII} must remain sufficiently negative.

The resistor R_{10} (fig. 5 and fig. 13) is the means of affecting U_{kIII} in such a way as to ensure the desired conditions. From fig. 13 it will be seen, assuming St_1 , St_2 to have ignited (their working voltage being 85 V), but not St_3 , that:

$$-U_{kIII} = 170 + \frac{R_{10}}{R_6 + R_{10}} (V_0 - 170).$$

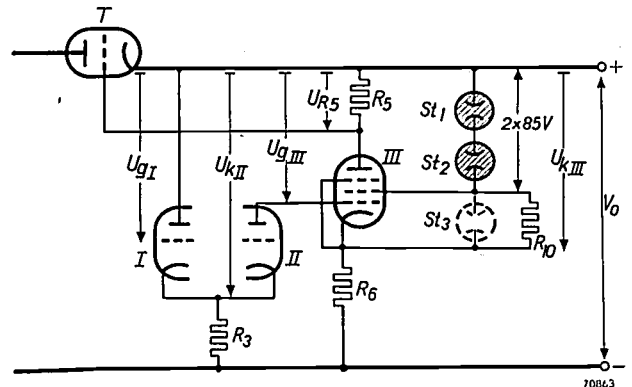


Fig. 13. In this diagram it is assumed that V_0 has risen to the value where St_1 and St_2 have ignited, but not St_3 . Potentials U_{gI} , U_{kII} , U_{kIII} , U_{gIII} and U_{R5} are all negative with respect to the positive line. These are shown as functions of V_0 in fig. 12. Other references as in fig. 5.

A reduction in the value of R_{10} therefore results in a drop in $-U_{kIII}$, i.e. an increase in the potential of the cathode of valve III. By making R_{10} small enough we can ensure that this potential will be maintained at a value that is sufficiently higher than that of the control grid.

A lower limit is set for R_{10} by the condition that in the stable state the current passed by the tube St_3 must not be too small. For efficient stabilisation a current of at least 1 mA should flow in this tube. With a further 1 mA cathode current in the pentode, there remains of the current flowing in R_6 (which is 4.45 mA in the stable state) at most 2.45 mA for R_{10} . It thus follows that with a working voltage of 85 V for St_3 :

$$R_{10} \geq \frac{85}{2.45 \times 10^{-3}} = 35\ 000\ \Omega.$$

Once this condition is satisfied, another risk entailed by a reduction in R_{10} is removed as well, viz. that when V_0 rises, the voltage across R_{10} might not be sufficiently high to ignite

St_3 . It is found that the minimum value of R_{10} necessary for infallible ignition of St_3 is 23 000 Ω .

The highest value that can be attributed to R_{10} without resulting in a premature flow of current in the pentode is 95 000 Ω . The actual value of R_{10} has thus been placed at 50 000 Ω .

Regarding the form of the curves U_{gI} , U_{kII} , U_{gIII} and U_{kIII} in fig. 12, the following may be added.

The potential U_{gI} of the grid of valve I (fig. 5) with respect to positive is $-V_B$. Current can flow in this valve only when the cathode potential U_{kII} approximates to the grid potential U_{gI} . As long as St_1 and St_2 have not ignited and valve II and resistor R_3 are thus not passing current, $U_{kII} = -V_0$ (U_{kIII} and U_{gIII} are then also $-V_0$, see fig. 12). When St_1 and St_2 ignite at $V_0 \approx 210$ V, U_{kII} is still so far removed from U_{gI} that I is still not conductive. But, once St_1 and St_2 have ignited, current commences to flow in II and R_3 . Because of the voltage drop across R_3 , U_{kII} now rises less quickly than V_0 , and it is only when V_0 is in the region of 400 V that I commences to work. From that point onwards U_{kII} remains constant at $\approx -V_B$.

In these circumstances the difference between the anode and cathode potentials of II does not exceed about 15 V and curve U_{gIII} (= anode voltage of II) therefore runs parallel with and close to U_{kII} .

When St_3 has ignited, $-U_{kIII}$ is constant at a value equal to the sum of the three working voltages, viz. 255 V.

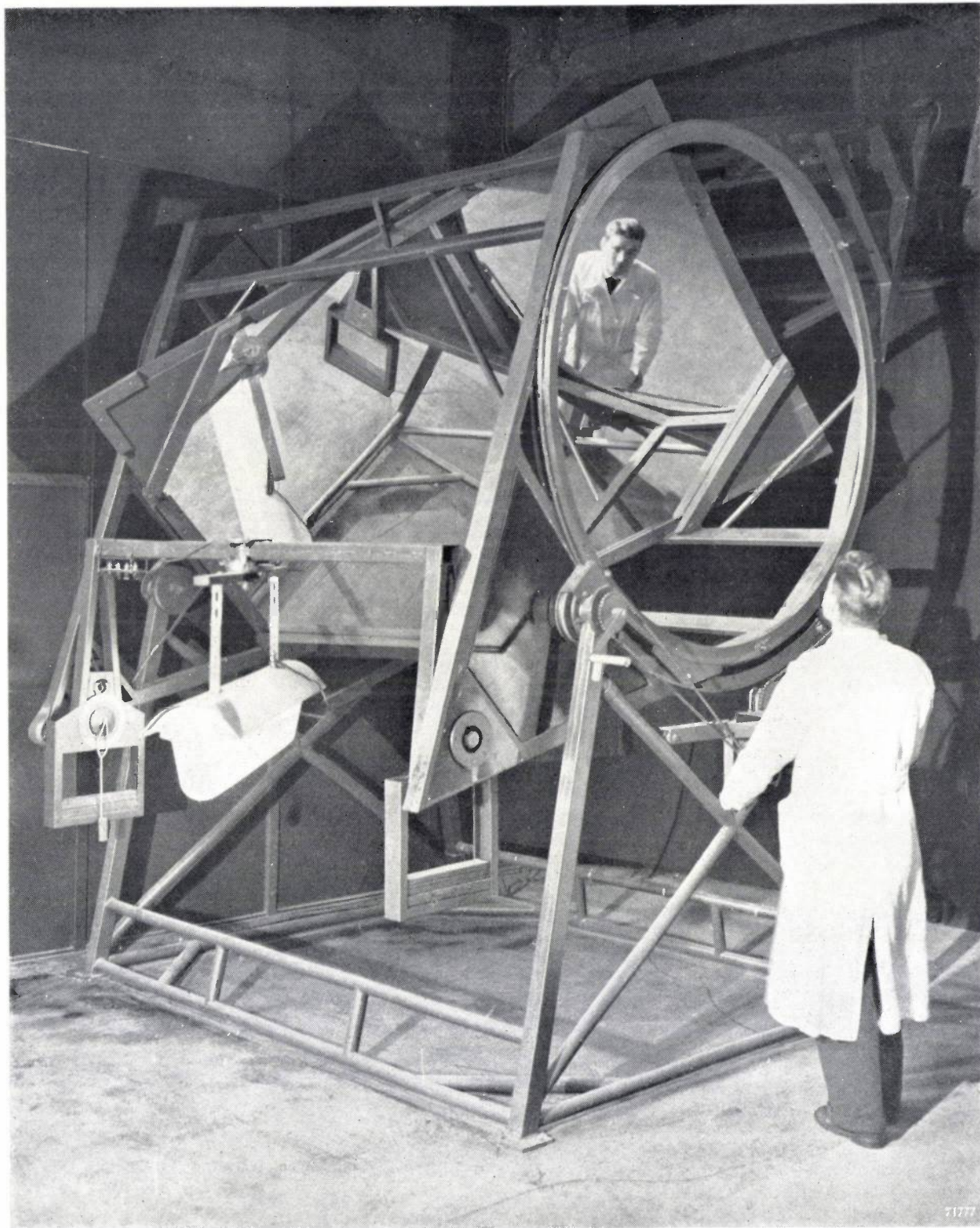
Summary. A rectifier has been designed for laboratory work in the field of decimetric and centimetric waves, to provide a stabilised voltage that will ensure a very constant frequency from transmitting valves fed from the rectifier. Relative fluctuations in the output voltage are only 1/3000th of those in the mains supply. The internal resistance is less than 20 Ω for H.F. currents. With a direct voltage output of 5000 V the ripple and noise do not exceed 10 mV.

This result is attained by employing a well-known method based on the use of a triode control valve which functions as a variable resistor in the D.C. circuit. A fraction of the output voltage is compared with a reference voltage and the difference is amplified in a direct voltage amplifier of which the output voltage is used as grid bias for the control valve. The circuit of this amplifier, which is fed from the high output voltage from the rectifier, is of an unusual type. It comprises two stages giving an overall gain of 3000, with an input stage giving no gain (cathode follower). The reference voltage is supplied by dry batteries (300 V). The output voltage is variable from 900 to 5100 V in four stages of 1200 V, 6 stages of 200 and continuously over a range of slightly more than 200 V. The high direct voltage and gain entail special conditions as regards assembly and quality of the various components.

Some of the precautionary measures taken with a view to safeguarding the equipment and preventing instability are discussed, as is also the method of preventing the output voltage from assuming too low a level of equilibrium.

APPARATUS FOR MEASURING LIGHT DISTRIBUTIONS

535.245.002.54

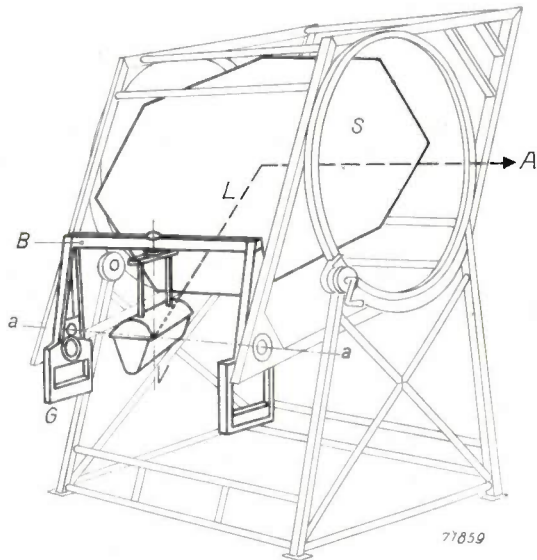


The apparatus shown in the reproduction was designed in the Light Technical Laboratory at Eindhoven for measuring the luminous intensity, in all directions, of light sources of all kinds and sizes. The intensity is measured by means of a photo-electric cell, which is set up on the right-hand

side at a large distance, in a fixed position; in the drawing on the opposite page the photo-electric cell is situated on the extended part of the dotted line *A*.

The apparatus consists mainly of a sort of cage which can be rotated about the above-mentioned

horizontal dotted line as its axis. In the cage a large flat mirror (*S*) is mounted at an angle of 45° to this axis. The mirror always reflects a beam of light from the light source to be examined,



emitted in the direction of the axis *L*, in the direction *A* to the photo-electric cell, independent of the position of the cage. The light source — in the photographs it is a fitting for street lighting with four fluorescent lamps (“TL” lamps) — is suspended in a bracket (*B*), which can turn freely about the secondary axis *a-a* passing through the centre of the light source. By means of weights (*G*) the bracket, and consequently the light source, are always kept in a vertical position when the

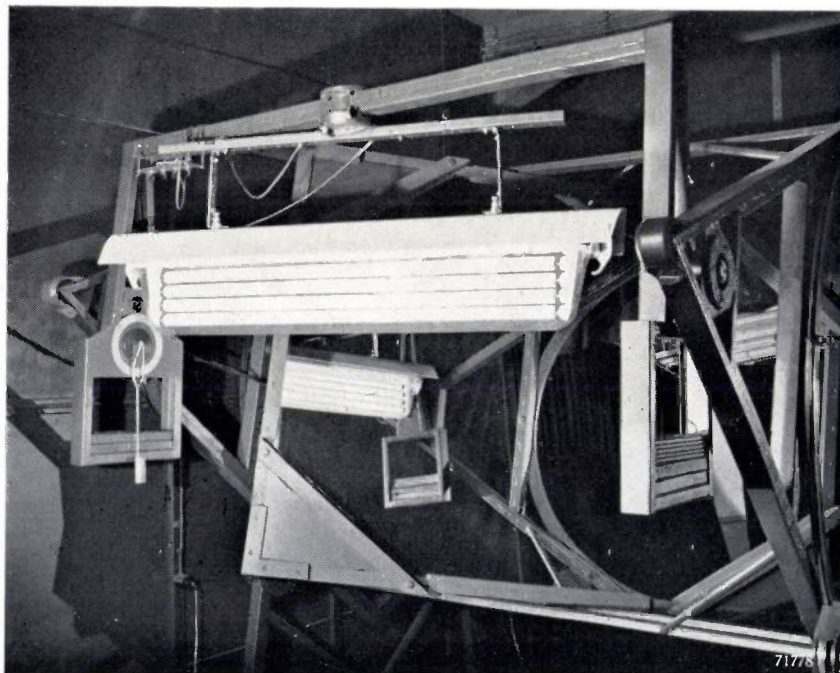
cage is turned. Therefore the photo-electric cell “scans” the emitted light in a meridian round the light source at each complete revolution of the cage.

By turning the light source through a small angle about the (always vertical!) axis *V* after each complete revolution, a new meridian can be scanned each time, and in this way the total light distribution can be determined.

The close-up below shows the suspension of the light source in the bracket in a different position. By turning the bracket upside down, the light distribution of standing fittings can also be measured.

Of course the construction would have been simpler if the light source itself had been placed with its centre on the axis *A* and was revolved about two axes perpendicular to *A*. However, this would involve tilting over the light source from its normal burning position; several types of lamps cannot be used in any position; with other types it would lead to inaccurate measuring results. For these reasons the apparatus would not be suitable for these lamps. Therefore, in other apparatus designed for the same purpose, the light source is set up in a fixed position, either the photo-electric cell itself or one or more mirrors, which lead the light from the source to a fixed photocell, swinging about the light source. The depicted apparatus compares favourably to such constructions in that, with relatively small dimensions of the apparatus, the photo-electric cell can be placed at a very large distance¹⁾

¹⁾ This is necessary for light sources with a large “photo-metrical boundary distance” (e.g. projectors); see Philips techn. Rev. 9, 114, 1947.



and that no excentric movement of a large, heavy mirror is necessary. In our case the acting forces and couples are mainly symmetrical and relatively small, so that inadmissible inaccuracies in the setting of the measuring direction due to all sorts of deformation are avoided with a relatively light and simple construction.

The last-mentioned point of view has also lead to the peculiar cage construction. Due to the axially emitted beam of light, the rotating construction on this side could not be supported by a normal bearing. With one single bearing on the other side the required rigidity could only have been obtained at the cost of a very heavy and expensive construction. The solution shown in the figures with two large rings each resting on two roller bearings in a pedestal is light and yet rigid. As a consequence

of the bearings being fitted on both sides, we only observe vertical sagging at the suspension points of the fitting and mirror and no angular deformations, which are much more objectionable if accuracy of measurement is required. The sagging of the bracket where the fitting is suspended is less than 2 mm with a fitting weighing 100 kg, and, moreover, of no consequence, as it is automatically taken into account during the adjustment of the fitting. The largest observed sagging of the mirror is less than 0.1 mm. Consequently the angle measurements are most reliable: in every meridian the angle ("elevation") can be adjusted to $1/20^\circ$, the angle of the meridian itself ("azimuth") to $1/12^\circ$. Accuracies of this order are required, among others, for some lights for car and airfield lighting.

L. de WIT.

HARDENING OF METALS

by J. L. MEIJERING.

539.53:621.785.61

For many technical applications a metal having a high degree of hardness is required. This hardness is obtained by alloying metals which are not hard enough in their pure condition with other elements: sometimes some sort of heat treatment or a deformation process follows. The most important hardening methods are quenching (steel hardening) and precipitation hardening. A new method was developed in the Philips Laboratory at Eindhoven, viz. the so-called oxidation hardening. This method, which is still in the experimental stage, is discussed briefly in this paper. We have availed ourselves of the opportunity to discourse at some length on the hardness of metals in general and on the physical phenomena which form the basis of the hardening process in particular.

Hardness

The hardness of a material in the widest sense of the word is understood to be the resistance which must be overcome to deform this material plastically, i.e. permanently. The resistance against elastic deformation, i.e. deformation which returns automatically to zero after removal of the deforming forces, is governed by the elastic properties of the material. We shall not discuss this subject here.

In a narrower sense hardness may be defined as the resistance against impression. This hardness is often expressed in so-called Vickers units. In quite a number of cases "hard" may also mean: "of great tensile strength". Generally speaking, a fair correlation is found between tensile strength and Vickers hardness for a given material.

The crystal structure is of importance with regard to the plastic properties: as the number of suitable glide planes in the crystal is larger, so the malleability will be greater. Thus the great hardness and brittleness of silicon and germanium is attributable to the special structure of these elements, viz. the so-called diamond structure, in which each atom has four adjacent neighbours (fig. 1d) and in which no proper glide planes occur. The occurrence of the diamond structure (the prototype of which, viz. diamond, is extremely hard) in silicon and germanium is closely connected with the homopolar character of these "semi-metals".

These semi-metals will not be discussed here; we shall turn our attention to the body-centred cubic

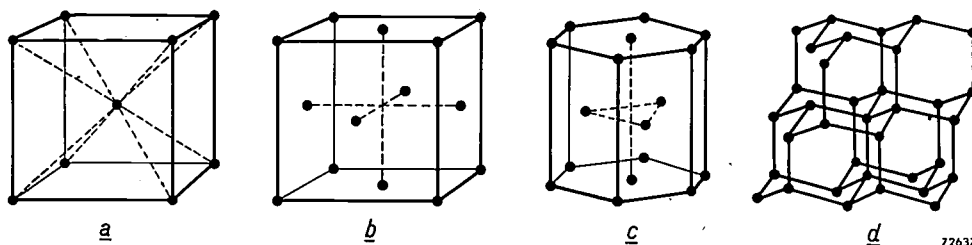


Fig. 1. The three structures in which most metals crystallise: a) body-centred cubic structure, b) face-centred cubic structure, c) hexagonal close-packed structure. The diamond structure is also depicted (d). The atoms have been drawn relatively much too small in these figures; they should touch each other.

In plastic deformation layers of atoms move along "glide planes" in the crystal. These glide planes are often those layers of atoms for which the distance between two successive parallel lattice planes is a maximum, and in which the packing in the plane (layer of atoms) itself is densest. Fig. 1 depicts a few crystal structures of frequent occurrence. In face-centred cubic crystallised metals e.g. (fig. 1b) the planes of closest packing are the (111)-planes.

arrangement (fig. 1a), to the face-centred cubic and to the hexagonal close-packed metals. The majority of the real metals has one (or more, in elements showing allotropy) of these three structures¹⁾.

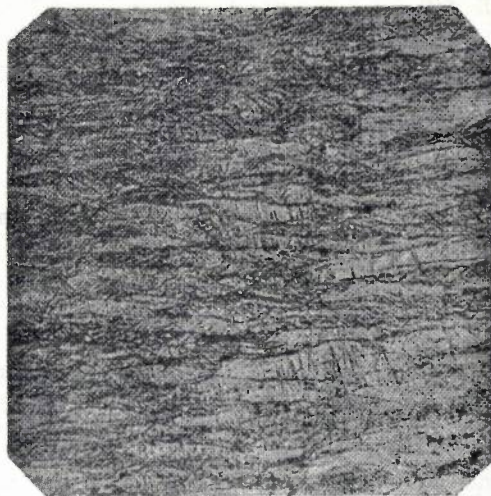
¹⁾ We wish to observe that a true metal such as manganese is also very hard and brittle if it occurs in the α or β modification, since the crystal structures of α and β manganese are complicated and have no well-defined glide planes. On the other hand, γ manganese, which is stable at high temperatures has a face-centred structure and is ductile.

The hardness of a metal in its pure condition is often not sufficient for technical purposes. In the majority of cases attempts are made to increase the hardness in one way or other, although for some applications it may be desirable that the metal is not hard. Generally speaking it is not possible to decrease the hardness permanently; on the other hand, various methods exist for increasing the hardness, which under the circumstances is a great advantage.

Hardness of pure metals

On comparing various pure metals of the same crystal structure, the hardness will (at a given temperature) be greater as the melting point of the

metal at a given temperature may be varied considerably, as mentioned before, chiefly in an increasing sense. The hardness increases by plastic deformation (rolling, drawing, hammering): this work-hardening is based on the fact that the lattice is distorted by the gliding of the atoms along the glide planes²⁾. It stands to reason that this method of hardening by rolling etc. can be realised with ductile metals only. The work-hardened material can be resoftened by heating: if the temperature does not rise too high and the material is not too strongly deformed, recovery may occur and the lattice distortions rearrange themselves in such a way that the total distortion is diminished. On heating to a higher tem-



a



b

Fig. 2. Recrystallisation of brass with 20% zinc; a) cold, rolled, b) subsequently annealed (taken from P. Goerens, Einführung in die Metallographie, Wilhelm Knapp, Halle 1932).

metal is higher. This is illustrated by the series Pb, Al, Cu, Ni, Pt. This is not surprising: a lower melting point means a weaker cohesion between the atoms. The hardness of one and the same metal decreases with rising temperature: the stronger thermal agitation involves an easier surmounting of the cohesive forces in the gliding process.

We wish to observe, in order to avoid misapprehension, that although greater hardness is, broadly speaking, often accompanied by greater brittleness (smaller elongation), this need not always be the case, even with one and the same metal. With the above-mentioned face-centred cubic metals, both hardness and elongation increase with decreasing temperature,

perature or after stronger deformation, recrystallisation occurs: a number of nuclei grow to new, undistorted crystals with orientations differing from those of the deformed old crystals which they consume (fig. 2). Consequently, the hardness decreases both on recovery and with recrystallisation. If the plastic deformation takes place at a temperature higher than the "recrystallisation temperature" (which naturally varies widely for different

²⁾ We may refer the reader to J. D. Fast, Strain ageing in iron and steel, Philips techn. Rev. 14, 60-67, 1952 (No. 2), for a detailed picture of these irregularities (dislocations).

metals, and also for a given metal depends greatly on the specific deformation process and the annealing treatment) no appreciable work-hardening will occur (hot rolling). With a metal having a low melting point the recrystallisation temperature may even lie below room temperature (as with lead).

The hardness of a completely recrystallised ("soft annealed") metal still depends more or less on the grain size. A grain boundary means a change in the orientation of the lattice and consequently in the direction of the glide planes, and therefore interferes with the deformation process. For this reason, very fine crystalline copper e.g. will be harder than a Cu single crystal.

At high temperatures, where creep occurs, i.e. a deformation that increases with time when a constant load is applied, the reverse is observed: a Cu single crystal is found to creep much slower than polycrystalline copper. Similarly, well annealed copper does not creep as quickly as copper that is less completely recrystallised after deformation. Apparently at these high temperatures the lattice distortions promote the deformation process. We may attempt to explain this by suggesting that the disappearance of all types of lattice faults, while heating under stress, causes creep. Experiments in this field have not yet given a decisive answer to this phenomenon.

A very much distorted lattice of a metal may also be obtained by building up the lattice in circumstances (e.g. at a rather low temperature) in which many lattice faults will occur. Very hard nickel³⁾ may, for example, be obtained by electrolytical deposition under special conditions. The presence of chemical impurities, in very small concentrations, chiefly dissolved gasses, e.g. hydrogen, is probably also of importance.

Hardness of alloys

For simplicity's sake we shall confine ourselves in the following to binary alloys. Alloys with three or more components show no essentially new points.

There are two main types of solid solutions: substitutional solid solutions, e.g. Fe-Ni, and interstitial solid solutions, e.g. Fe-C. In the first case, with which we are nearly always concerned if both components are metals, Fe atoms are displaced by Ni atoms on, in most cases irregularly distributed, lattice points. In the second case — which occurs if the added element is a non-metal having atoms which are small in comparison to the atoms of the parent metal — the foreign atoms are built in at

places (interstices) between the lattice sites of the parent lattice.

We may expect the hardness of the solid solution to be greater than that of the parent metal (the latter must be understood to mean the metal with the greatest concentration), since interstitial atoms will disturb the order of the lattice, especially if they are larger than the normal interstices in the lattice of the pure parent metal. Substituted atoms will also deform the lattice, proportionately more as the atoms of the metal added to the alloy differ more in size from those of the parent metal. In a dilute solid solution the hardness increase is approximately proportional to the concentration in foreign atoms, and according to a rule of Norbury, when various metals are added, the change in hardness per atomic per cent addition in a given parent metal (e.g. copper) is approximately proportional to the absolute value of the difference in atomic diameters, as determined from the lattice constants of the pure metal⁴⁾.

It is only an approximation that a metal lattice is composed solely of (e.g. spheric) atoms. The valency electrons are more or less detached from the ions on the lattice points and together form the "electron gas", which is responsible for a high electrical and thermal conductivity. For this reason Norbury's rule should not be taken too literally, especially if the differences in atomic diameter are small. Silver and gold e.g. have practically the same atomic diameter (the same face-centred cubic structure with a difference in lattice constant of only 2^{0/100}), but still silver is appreciably hardened by alloying it with 5 at. % gold (and vice versa). It is, however, a fact that the hardness increases more quickly with the concentration if the differences in atomic diameter are large than if they are small.

Viewed from this angle, the question: "Given a certain metal, in what manner can it be hardened as much as possible by alloying it with e.g. 5 at. % of another element?" should be answered as follows: "Add an element the atoms of which differ as much as possible in size from the parent metal, e.g. alloy copper with lead". One circumstance, however, prevents us from acting on this principle. For the substitution of very large (or very small) atoms in a metal lattice costs so much energy, that the solubility becomes very small and the greater part of

⁴⁾ A. L. Norbury, *Trans. Faraday Soc.* 19, 586, 1924. M. A. Meyer (Diss. Delft 1951) also found this for solid solutions of nickel. Other authors found a hardening proportional to the square of the "misfit", for Cu and Ag solid solutions (see Diss. Meyer, page 57); Mott and Nabarro ("Strength of solids", Bristol Report 1948) also found this theoretically.

³⁾ The same applies to silver, see E. Raub, *Z. Metallk.* 38, 87, 1947, and to chromium.

the metal added to the alloy occurs in the form of a second phase rich in this metal; the homogeneous alloy changes into a heterogeneous one.

Heterogeneous alloys

The hardness of an alloy consisting of two phases is, roughly speaking, a volume average of the hardnesses of the two phases, measured as micro hardnesses *in situ*⁵⁾. In most cases, and nearly always

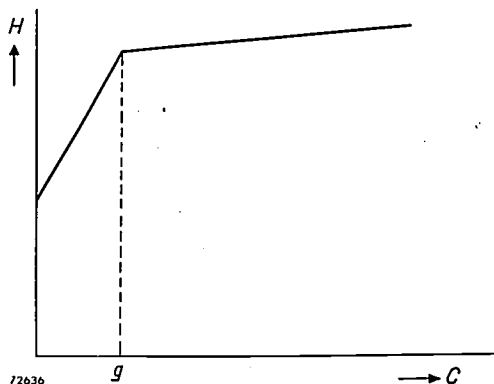


Fig. 3. The hardness H of a binary alloy increases as the concentration C of the alloy component in the parent metal increases, and that approximately proportional to this concentration, as long as the alloy component can still be dissolved in the parent metal (homogeneous alloy). If the saturation point g is passed, however, and the second phase is being formed, the alloy becoming heterogeneous, the change in hardness chiefly depends on the hardness of the new phase. The increase in hardness becomes much slower and sometimes even changes into a decrease.

if the concentration of the element added to the alloy is relatively small, the total hardness develops as shown in *fig. 3*. Consequently, the hardness is smaller in the heterogeneous field after passing the saturation point than it would be if the solution was supersaturated homogeneous (as shown by extrapolation); the latter hardness can in some cases be experimentally determined in supersaturated solutions. If the atomic diameters of both elements differ more than 15 to 20 %, the saturation point (g in *fig. 3*) is always very low, even at high temperatures (Hume-Rothery)⁶⁾. In these cases, according to Norbury, the hardness will rise steeply in the homogeneous field, but the attained increase in hardness at the saturation point is still negligible because the homogeneous field is so narrow. After passing the saturation point, the hardness rises less

⁵⁾ Roughly, because analogous to the influence of the crystal size in homogeneous alloys, the hardness of a heterogeneous alloy will be greater as the second phase is more finely divided.

⁶⁾ Although a small difference in atomic diameter is required for a proper solubility, it does not guarantee this. The solubility of Fe in solid Cu for example, is only very small, although Fe and Cu differ only 1% in atom diameter (and intermetallic compounds are not formed).

steeply, or in some cases even falls, as e.g. in the system Cu-Pb, where the second phase, nearly pure lead, is softer than the parent phase.

Hardening methods

It is possible to bring large atoms (or molecules) into the lattice and so harden the metal, albeit that the resulting alloy is thermodynamically not stable⁷⁾. This is very important for practical hardening methods.

The chief hardening methods are those applicable after the forming, e.g. rolling or drawing of the metals, and which result in a greater hardness of the material than it possessed during the forming operation. We may distinguish three methods: quenching from the solid phase (martensitic hardening), precipitation hardening and oxidation hardening, which will be discussed in succession. The first two methods have been known for a long time already, we shall therefore only discuss them briefly. The third method, the oxidation hardening, was developed in the Philips Laboratory at Eindhoven during the last decade: this method will be discussed somewhat more extensively in this paper.

A relatively hard metal may also be obtained by a special treatment before or during the forming operation. One of these methods is quenching liquid alloys⁸⁾. In molten copper as much as 10 at. % of lead is soluble, and it is conceivable that, provided the cooling is effected quickly enough, a considerably larger concentration of lead in the solid solution may be obtained than the 0.001 to 0.01% soluble in copper in the solid phase. The desired result might also be obtained both by quick condensation from the vapour phase of two metals at the same time, and by simultaneous electrolytical deposition.

There are not many data on the two first-mentioned methods. The electrolytical method is more important for practical applications, with this understanding however that very great hardnesses are also obtainable by electrolytical deposition in a special manner of e.g. nickel alone (see above)⁹⁾.

One of the processes by which the hardness of a metal is altered during the forming operation is the work-hardening already mentioned above. We shall not enlarge on this subject here either, but we do wish to point out in passing the peculiarities that may occur during the work-hardening of alloys. A most remarkable case occurs in the system aluminium-zinc. Zinc with approximately 20% aluminium can in some circumstances be much softer than zinc and aluminium themselves¹⁰⁾.

⁷⁾ Cold-worked metals and alloys do not possess this quality either, nor strictly speaking do poly-crystalline metals and alloys (surface tension!).

⁸⁾ For this subject see e.g. H. Bückle, *Z. Metallk.* **43**, 82 1952.

⁹⁾ Raub and Engel (*Z. Metallk.* **41**, 487, 1950) obtained strongly supersaturated solid solutions of Cu with Pb, and of Ag with Bi electrolytically. A great hardness was indeed obtained.

¹⁰⁾ A. A. Bocharov and Z. A. Sviderskaya, *Bull. Acad. Sci. U.R.S.S. Classe sci. techn.* 1945, 821; see also *Chem. Abstr.* **41**, 2375, 1947. F. Sauerwald *Arch. Metallk.* **3**, 165-173, 1949, studied this alloy independently: the elongation was found in some cases to amount to 180%.

Until now it is a unique thing for a binary alloy to be softer than its two components. This is a case of non-equilibrium conditions: the adjustment to equilibrium is promoted by the deformation and conversely the deformation is favoured by the fact that the atoms are rearranging themselves.

Martensitic hardening

In martensitic hardening we try to take advantage from the transition of a solid solution of one solid phase into a phase of a different structure, e.g. of the transition face-centred cubic into body-centred cubic, which may occur when cooling various iron alloys (austenite-ferrite). This method will not prove very successful in the case of substitutional solid solutions: if the atoms differ too much in size, the solubility is small for both structures. With interstitial solid solutions the case is different. In certain circumstances the deformation of the lattice will increase roughly proportionally with $(r-r_h)/r_h$ ¹¹, in which r is the atom radius of the dissolved element and r_h the radius of the inscribed sphere of an interstice in the (undistorted) parent lattice. It is possible for r_h to differ considerably for both structures, e.g. it can be larger for the phase stable at high temperature than for the phase stable at low temperature. In the first-mentioned phase the solubility for too large an atom will then be much higher.

On quenching, from e.g. 900 °C, of γ -iron (austenite) with e.g. 4 at. % (0.9% by weight) carbon dissolved in it, α -iron (ferrite) is obtained in which carbon is found in supersaturated solution (martensite). The strong hardening occurring in this case (i.e. the classical steel hardening that has already been applied for centuries) may be ascribed to the strong deformation caused by C atoms occupying too narrow places in the ferrite lattice.

We may roughly calculate the difference in lattice deformation by an interstitial atom in the body-centred or face-centred cubic structures respectively as follows. Assume the lattice deformation per atom dissolved to be proportional to $n(\Delta r)^2$. In this formula n stands for the number of adjacent neighbours per atom: 12 for a face-centred cubic substitutional solid solution, 6 for a similar interstitial solid solution, 8 for a body-centred cubic substitutional solid solution, 2 or 4 for a similar interstitial solid solution. Δr stands for $r-R$ in a substitutional solid solution, viz. the difference in atomic radii of the dissolved metal (r) and the parent metal (R); in an interstitial solid solution Δr stands for the difference between the atomic radius of the foreign atom and the radius r_h of the interstices in the parent metal. We have assumed the exponent of Δr to be 2, which follows from purely geometrical considerations for $n(\Delta r)^2$ is proportional to the volume occupied "doubly", if Δr is small, assuming the positions and radii of the atoms present in the metal to remain unchanged

¹¹ If $r > r_h$, which is normally the case.

when the foreign atom is brought into it. An analysis of the increase in hardness proportional to the concentration of homogeneous alloys (Norbury's rule) shows that actually the exponent of Δr lies probably between 1 and 2 (compare note ⁴).

Let us consider the face-centred cubic structure. An interstice has only half as many adjacent neighbours as an atom has. Moreover, according to elementary mathematics, $r_h = R(\sqrt{2}-1)$. When discussing the heterogeneous alloys we saw that, generally speaking, a difference in atomic radius of 15% is still only just tolerable for a substitutional solution to be formed. According to our formula, a ratio $r/R = 1 + u$ for substitutional solid solutions corresponds, as regards lattice deformation, to a ratio $r/R = \sqrt{2}-1 + u\sqrt{2}$ for interstitial solid solutions. If we substitute the value 15% for u , we find that the addition of atoms with a diameter of approximately 63% of the diameter of the atoms of the parent metal would be just tolerable for an interstitial solution to be formed.

In the body-centred cubic structure two different kinds of interstices are suitable for occupation¹². The largest interstices have four adjacent neighbours, and in this case $r_h = R\left(\sqrt{\frac{5}{3}}-1\right)$; the smaller interstices have two adjacent neighbours, and $r_h = R\left(\sqrt{\frac{4}{3}}-1\right)$. In this case we find that $r/R = 1 + u$ for substitutional solid solutions corresponds to $r/R = \sqrt{\frac{4}{3}}-1 + 2u$ for interstitial solution in the smallest interstices, and to $r/R = \sqrt{\frac{5}{3}}-1 + u\sqrt{2}$ for interstitial solution in the larger interstices. Substitution of $u = 0,15$ gives 0,455 R and 0,50 R respectively for the largest atom diameter tolerable for interstitial solution to be formed.

We shall now consider the steel hardening somewhat closer. The radii of C atoms and N atoms amount to 61% and 55% of the radius of the iron atom respectively. This explains why carbon and nitrogen dissolve rather well in austenite (face-centred structure), and only to a very low degree in ferrite (body-centred structure); a transition of austenite into ferrite on quenching should give a strong deformation and consequently great hardening.

We shall have to bear in mind that the ratio of the atomic radii is not the only factor determining the solubility; however, we shall not discuss the other factors in this paper.

Precipitation hardening

A metal lattice will also be considerably deformed if we succeed in producing groups of atoms or molecules in the lattice in one way or other. Precipitation hardening is an example of a hardening method based on the first-mentioned process. Precipitation hardening (as also martensitic hardening) is often applied in practice; the following may serve as an example. Aluminium atoms and copper atoms differ $\pm 10\%$ in atom radius. At room temperature the solubility of copper in aluminium is practically zero, but at 550 °C (at the eutectic temperature) 2 at. % Cu are soluble in aluminium. After quenching from this temperature to room

¹² See J. D. Fast, Ageing phenomena in iron and steel after rapid cooling, Philips techn. Rev. 13, 165-171, 1951 (No. 6).

temperature the solid solution is supersaturated. The hardness is normal: somewhat harder than in the stable state, in which practically all copper is precipitated in the form of coarse CuAl_2 particles as a second phase. This stable final state can never be attained at room temperature, since the diffusion rate is much too low. When annealing at temperatures of roughly 300 °C (where the solubility is still low), the final state is approached much closer within a measurable space of time. It is of importance that this final state is reached via intermediate states, in which the hardness is considerably greater than in the initial and final states; there is thus a maximum hardness. These intermediate states, which are characterised by the presence of knots of Cu-atoms which deform the lattice (see the article referred to in note ¹²) can be fixed at room temperature by cooling. By a proper duration and temperature of the annealing treatment, very strongly hardened materials can be obtained in this way. A considerable hardness is particularly obtained by "annealing" at temperatures below approximately 200 °C. In proportion as the temperature is lower, it takes more time to reach the optimum. Hardening also occurs at room temperature, but the maximum is not attained.

We have already observed before in this article that a heterogeneous alloy is harder in proportion as the second phase is more finely divided. In the above-mentioned state of maximum hardness the particles are not visible under the optical microscope; this is a case of transition homogeneous-heterogeneous. The (vague) boundary between homogeneous and heterogeneous states apparently lies at various average particle sizes, judging from various properties.

The existence of such a thing as an optimum size of the particles may be explained by the following purely geometrical reasons, to which no quantitative value should be attributed however. Let x be the radius of the particles assumed to be spheric, then the concentration of the particles is inversely proportional to x^3 . We have seen that if $(x-R)$ is small, the deformation of the lattice resulting from the presence of the particles increases with $(x-R)^p$, where p lies between 1 and 2. Assuming this to hold good also for larger values of $(x-R)$, the total deformation is proportional to $(x-R)^p/x^3$, which expression gives a maximum for $x = 3R/(3-p)$. According to this reasoning we could expect particles of a few times the atomic radius of the parent metal to be most effective in precipitation hardening. However, the influence of internal stresses and surface tensions renders the quantitative result of the calculation very uncertain.

We wish to observe that in quite a few cases, for example in the system Al-Cu, the particles are not spherical but in the form of platelets. The smallest

dimension, in this case the thickness of the platelets, seems to amount to a few times the atomic radius.

It stands to reason that there will be a tendency towards such submicroscopical precipitation phenomena also in the case of steel hardening. In this case the stable final state consists of practically pure Fe plus coarse carbide particles Fe_3C (or, still more stable, graphite). There is every possibility in this case of an initial "coagulation" resulting in still greater hardnesses than the "ideal" state in which the carbon is divided atomically. Precipitation phenomena will also occur in the above-mentioned electrolytical alloys ¹³).

For a more profound discussion of steel hardening and precipitation hardening in general we may refer to the very extensive literature existing on the subject ¹⁴).

Oxidation hardening

Oxidation hardening, or rather hardening by internal oxidation, is a method of hardening based on the formation of foreign molecules in the crystal lattice of the basic metal.

We wish to discuss this method somewhat more extensively than the methods mentioned earlier in this paper. First of all we shall illustrate the oxidation hardening by an example. On heating in air silver in which 1.5 at. % magnesium has been dissolved (substitutionally), a perceptible quantity of oxygen diffuses into the metal. This oxygen combines with the Mg-atoms to form molecules of MgO , and these molecules deform the silver lattice considerably by their large dimensions. Consequently the metal becomes harder; the Vickers hardness increases from about 40 to approximately 170.

The above shows that oxidation hardening is caused by internal oxidation of a component of an alloy. The occurrence of internal oxidation has been known for a much longer time than the hardening phenomena that may accompany this oxidation ¹⁵). Internal oxidation will not in all cases proceed in the manner discussed above. One of the conditions for a process to occur in analogy with that in the example given above is that the oxidising component occurs in a relatively small concentration in the solid

¹³) Raub and Engel ⁹) indeed found a hardness optimum, when tempering their electrolytically deposited alloys, and the same is found when tempering martensite. (see K. H. Jack, *J. Iron Steel Inst.* **169**, 26, 1951).

¹⁴) We may refer the reader for a survey of the most important phenomena in steel hardening to F. Seitz, *The Physics of Metals*, chapt. XIV, Mc. Graw Hill, New York 1943. A survey of precipitation hardening is found in G. C. Smith, *Progress in Metall Physics* **1**, 163-234, 1949.

¹⁵) See e.g. C. S. Smith, *Min. and Met.* **11**, 213, 1930; **13**, 481, 1932; K. W. Fröhlich, *Z. Metallk.* **28**, 368, 1936; F. N. Rhines *Met. Technol.*, **7**, T.P. 1162, 1940.

solution, the parent material itself remaining unaffected, so that the metallic properties of the alloy are not lost.

We shall in the following discuss (dilute) binary alloys only. The conditions for the formation of an oxide of the dissolved metal in the metallic parent metal are:

- 1) The oxygen should diffuse more rapidly into the metal than the atoms of the dissolved metal diffuse to the outside.
- 2) The affinity of the dissolved metal for oxygen should be greater than the affinity of the parent metal for oxygen.

As to the first condition: oxygen diffuses rapidly through silver, copper, nickel, γ -iron, titanium and zirconium. The diffusion coefficients D_M of foreign metals dissolved in one of these parent metals are mutually roughly equal and much smaller than the diffusion coefficient D_O of oxygen (at the same temperature). It is practically impossible to comply with the second condition for internal oxidation in the case of titanium and zirconium. Consequently, the parent metals suitable for internal oxidation are silver, copper, nickel and iron.

The oxidation in the metal progresses from the outside to the inside. The penetration depth d consequently increases with time; a good approximation of the penetration depth as a function of time is given by:

$$d = \alpha \sqrt{t}.$$

This can also be deduced theoretically¹⁶). Fig. 4 gives an illustration of the penetration. The factor α contains the diffusion coefficients D_M and D_O of the



Fig. 4. Internally oxidized silver strip with 1.2 at. % magnesium. The strip was heated for 50 minutes in air at 800 °C and afterwards sectioned, polished and etched. The oxidation proceeding from the outside to the inside may be clearly seen: a "rim" covering about $\frac{1}{4}$ of the thickness of the strip is oxidized, the interior is untransformed.

dissolved metal and the oxygen in the parent metal respectively, and the concentrations C_M and C_O in which the dissolved metal and the oxygen occur in the metal (the latter concentration as measured at the metal surface). If the ratio $C_M D_M / C_O D_O$ is much smaller than unity (e.g. approximately 0.01),

a case which is practically important, d does not depend on D_M .

The concentration of the oxide molecules in the oxidized layer (for the present we can neglect the diffusion of these molecules) is, if M is a bivalent metal, given by

$$C_{ox} = C_M(1 + C_M D_M / C_O D_O),$$

and, consequently, is somewhat higher than C_M in the above-mentioned case. Naturally, the concentration cannot have this value everywhere in a completely internally oxidized metal, since it is im-

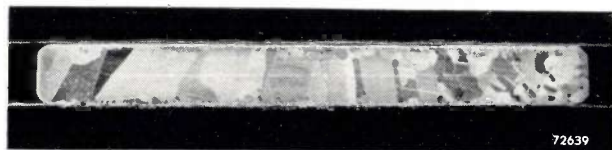


Fig. 5. In a strip completely oxidized internally (in this case silver with 1.5 at. % manganese) a "middle-line" poor in oxidized metal is left.

possible for the number of oxide molecules formed of the dissolved metal to be larger than the number of atoms of this metal. Therefore, if a thin metal strip is completely internally oxidized from both sides, we can expect a very thin zone in the middle of the metal where the oxide concentration is very much poorer. As a matter of fact such a "middle-line" is indeed found, as may be demonstrated by etching (fig. 5). Even after hundreds of hours of annealing the internally oxidized metal, this middle-line will still be found; this proves that the diffusion of oxide molecules is a very slow process indeed. Still it cannot be absolutely neglected, for the oxide molecules will dissociate to a certain extent into metal atoms and oxygen atoms, which do diffuse each by themselves through the metal and can recombine somewhere else in it. In this way coagulation of the oxide molecules to coarser particles can occur indirectly. The smaller the difference in affinity for oxygen between dissolved metal and parent metal, the easier will this take place. Since a coarse precipitate gives no proper hardening (see previous section), the terms of the second condition should be more strict in order to render oxidation hardening possible: the difference in affinity should be large (a quantitative thermodynamic discussion is found in¹⁶)).

We wish to point out here the difference between internal oxidation of silver alloys and internal oxidation of alloys of a less precious metal, such as copper. In the first case no oxide layer of the parent metal (silver) is formed at the surface of the metal. The oxygen concentration at the surface, C_O , is

¹⁶) J. L. Meijering and M. J. Druyvesteyn, Hardening of metals by internal oxidation, Philips Res. Rep. 2, 81-102, 1947 and 2, 260-280, 1947.

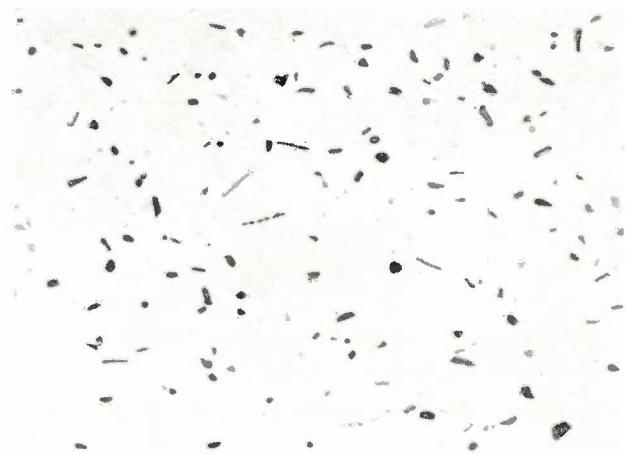
then directly proportional to the root of the oxygen pressure in the ambient atmosphere. The less precious metals are covered with an oxide layer (e.g. Cu_2O) at their surface on annealing in air: the internally oxidized metal only begins under this completely oxidized surface layer. In practice the resulting metal loss during the hardening process can be reduced by replacing the oxygen in the ambient atmosphere by nitrogen after a short time. The oxide already formed will supply sufficient oxygen for the completion of the internal oxidation within the metal.

We shall now give a few examples of the increase in hardness obtained by internal oxidation. By heating the silver alloy with 1,2 at.% magnesium in air for 2 hours at 800 °C, its Vickers hardness rose from the original value 40 to 170. Replacement of the magnesium by aluminium, beryllium and manganese gave final hardnesses of 160, 135 and 140 respectively (with practically the same concentrations of the alloys and equal annealing time). Zinc, tin and cadmium as components of the alloys caused no or only very little hardening. This will be readily understood: the oxides of these elements are not particularly stable.

Alloys of the composition $(100-x)\% \text{Cu } x\% M$, in which M stands for resp. Be, Al, Mg, Ti or Zr, and x lies around 1.5, showed an increase in hardness of appr. 40 to 165, 105, 105, 125 and 100 respectively. We found that manganese did not cause considerable oxidation hardening in copper, but it did in silver. This is due to the fact that the affinity of copper for oxygen is greater than that of silver for oxygen (more stringent condition 2). In this connection we wish to observe that it was only possible to bring about oxidation hardening in nickel (NiO is still more stable than Cu_2O), when aluminium was added. Oxidation hardening has not been found in iron alloys.

As said before, as in precipitation hardening, we shall also have to expect coagulation of oxide molecules occurring during the ageing of the alloy to affect the hardness, in the case of alloys hardened by internal oxidation. There may again be an optimum size of the particles giving the greatest hardness. We have not been able to verify this conclusively by experiments. It is beyond the scope of this article to discuss the problem of the coagulation in internally oxidized alloys; reference is made to an article published a few years ago¹⁷⁾. One can clearly observe oxide particles in internally oxidized Ag-Cd and Cu-Zn alloys (1 at. % alloy component) under

the optical microscope at a magnification of $500\times$ (fig. 6); in (1 at. %) Ag-Zn alloys this was not possible. In most cases the hardness has already decreased considerably before the coagulation has proceeded so far that the particles are microscopically visible.



72615

Fig. 6. An etched section through an internally oxidized Cu-Zn alloy (99% Cu-1% Zn) clearly shows oxide particles, when enlarged 500 times. The material was heated for $3\frac{1}{2}$ hours at 950 °C in air.

A property of alloys which is very sensitive to coagulation is the electrical resistivity. In the case of alloys hardened by precipitation, the resistance mostly rises first with increasing coagulation, after which it falls below the original value. The resistivity maximum is reached sooner than the hardness maximum. In some cases we found an electrical resistivity for internally oxidized silver alloys which was considerably larger than for non-oxidized alloys; in the case of copper alloys hardened by oxidation, no increase in resistance is found. This shows that the coagulation in copper alloys proceeds much further than in silver alloys because Cu has more affinity for O than Ag.

The coagulation also causes the deformation of "inner oxide films" in some alloys containing rather strong concentrations of base elements (fig. 7). These more or less coherent layers of oxide, roughly parallel to the metal surface, may impede the internal oxidation in the deeper parts of the metal¹⁸⁾.

The hardness obtained by oxidation hardening is less sensitive to subsequent treatment at high temperatures than that obtained by cold working or precipitation hardening, which is a definite advantage. The occurrence of intercrystalline brittleness on the other hand is a disadvantage.

¹⁷⁾ J. L. Meijering, Hardening of metals by internal oxidation, "Strength of Solids", Bristol Report, 140-150, 1948.

¹⁸⁾ J. L. Meijering, Pittsburgh International Conference on surface reactions, 103, 1948.



Fig. 7. Inner oxide films formed in copper with 5 at. % beryllium. The films run roughly parallel to the metal surface, the arrow indicates the direction of the oxygen diffusion.

Apart from hardening by internal oxidation, hardening by internal nitriding, viz. nitride

hardening of steel is known. If steel with a small content of e.g. aluminium or chromium is heated at approximately 500 °C in ammonia gas, a very strong increase in hardness is observed. This phenomenon may be explained by analogy with the picture given above of oxidation hardness, except that nitrides (of Al or Cr) will occur instead of oxides.

Summary. The hardness of a metal is its resistance against plastic deformation. This resistance may vary widely for the various kinds of metal; some factors influencing it are the crystal structure (both distorted and undistorted), the presence of foreign atoms in the lattice, or the presence of foreign molecules or groups of atoms. The hardness of a metal may be affected in various ways: by heat treatment either with or without chemical reactions, by deformation (work-hardening), or by a special method of producing the metal (electrolytical deposition). Thermal methods of hardening after the forming, such as quench hardening, precipitation hardening and oxidation hardening, are discussed and these methods are illustrated by examples, the last, new, method somewhat more extensively.

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.

2005: L. A. Ae. Sluyterman: Reactions of polypeptide esters in the solid state (thesis; Centen, Amsterdam 1951; 36 pp., 9 figs.).

Discussion of the behaviour of the methyl esters of the tripeptides diglycyl-glycine and alanyl-glycyl-glycine and of the tetrapeptides triglycyl-glycine and alenyl-diglycyl-glycine. It is shown that the tripeptides and the tetrapeptides react differently on heating in the solid state. The products obtained on heating have been investigated in a number of cases. In the discussion attention is drawn to the fact that the occurrence of an induction period and the migration of methyl groups as observed in some cases, and which preponderates above polycondensation, may be thought to be due to the rigidity of the crystal lattice.

2006: H. Bruining: Opneembuizen voor televisie, I. Indeling en wijze van werken van opneembuizen (T. Ned. Radiogenootschap 16, 209-225, 1951, No. 5). (Television camera tubes, I. Types and mode of operation; in Dutch).

In this paper a general outline is given on different types of television camera tubes. After an introduc-

tion on the two ways of potential stabilisation of a surface of insulating material by scanning electrons of low and high velocity respectively, the constructions of the orthicon, image-orthicon, vidicon, iconoscope and image-iconoscope are discussed. The mechanisms of storage, measuring and wiping out of electric charges due to bombardment by photo-electrons or due to photoconductivity are treated in detail. At the end a survey is given of tubes with their special names (see Philips tech. Rev. 13, 119-133, 1951, No. 5).

2007: P. Schagen: Opneembuizen voor televisie. II. Eigenschappen en onderlinge vergelijking van de opneembuizen (T. Ned. Radiogenootschap 16, 227-242, 1951, No. 5). (Television camera tubes, II. Properties and inter-comparison of camera tubes; in Dutch).

A comparison is given of the performances of the four most important types of camera tube. The special tubes considered are: Tubes with low-velocity target stabilization: Orthicon (C.P.S.-emitter). Image orthicon (American type 5826). Tubes with high-velocity target stabilization: Iconoscope,

Image iconoscope (Philips type 5854). The following characteristics are discussed: (1) Characteristics connected with the method of target stabilization employed: a) linearity and gamma, b) effective exposure-time, c) spurious-signal generation, d) simplicity of installation and operation. (2) Some other characteristics: a) Signal-to-noise ratio, b) spectral response, c) stability, d) depth of focus, e) possibility of black-level restoration. (3) Sensitivity: for normal broadcasting purposes, where a good signal-to-noise ratio and reasonable depth of focus are required, the sensitivities of the tubes are found to be related as follows: Iconoscope: Image iconoscope: Orthicon: Image orthicon = 1 : 75 : 25 : 325 (see No. 2006).

R 177: W. Ch. van Geel and B. C. Bouma: La déformation des redresseurs électrolytiques et de quelques redresseurs à couche d'arrêt (Philips Res. Rep. 6, 401-424, 1951, No. 6). (The de-formation of electrolytic rectifiers and of some barrier layer rectifiers; in French.)

Electrolytic rectifiers and some other kinds of rectifiers can lose their rectifying properties by a de-formation process. By "deformation" we mean the act of sending a current through the system in the direction opposite to the current used during the formation of the oxide layer (blocking-layer). During the deformation the current passes in the direction of easy transmission. After deformation the system can be restored by sending a current through the system in the direction that was originally the blocking direction (re-forming). After re-forming the system recovers its rectifying properties. For re-forming also alternating voltage can be used. The rectifying properties can also be taken away and given back to synthetic rectifiers, e.g. Zr-ZrO₂-(CuJ+J). The same can be done with rectifiers with a resin barrier layer, where a resin layer is brought onto a semiconductor and covered with a metal alloy and also with the ordinary type of selenium rectifier, where an alloy of low melting-point is brought directly onto the selenium layer. It is suggested that in all these cases rectification occurs because the barrier layer is built up in two parts which owe their conductivity to charge carriers of different sign (electron and hole semiconductors). By deformation one of the part loses its own character of conductivity

and there remain two parts with the same sign of conductivity and therefore rectification disappears.

R 178: Th. P. J. Botden: Transfer and transport of energy by resonance processes in luminescent solids (Philips Res. Rep. 6, 425-473, 1951, No. 6).

Transfer and transport of energy have been investigated in solids systems in which excitation does not produce free electrons and holes, so that these processes, if they do take place, are brought about by resonance. A distinction is made between those systems in which transport of energy may occur via the base lattice (described in part I) and the so-called sensitized systems in which transport may occur only via particular centres of the lattice (described in part II). Part I deals with the investigations on alkaline-earth tungstates and molybdates activated with samarium. Upon excitation in the absorption band of the base lattice with ultraviolet radiation both the tungstate or molybdate and the samarium fluorescence are emitted at low temperatures (below about 250 °K). At these temperatures the quantum efficiency of the samarium fluorescence is independent of temperature while for excitation by certain wavelengths the quantum efficiency of the tungstate or molybdate fluorescence is temperature-dependent, because the absorption changes with temperature. The excitation energy that leads to samarium fluorescence at low temperatures is mainly absorbed in tungstate or molybdate groups in the neighbourhood of the samarium ions and the energy is transferred from these groups to samarium. At the temperature at which the tungstate or molybdate fluorescence is quenched, or at a slightly higher temperature, the quantum efficiency of the samarium fluorescence increases with temperature due to the transfer of energy from tungstate or molybdate groups at larger distances from the samarium ions to these ions. This energy is transported from the groups first excited to groups in the neighbourhood of the samarium ions. The mechanisms of transfer and transport of energy are discussed. Calculations have been carried out on the basis of the theories of Mott-Seitz and of Möglich-Rompe for the quenching of fluorescence. The results are in favour of the theory of Mott-Seitz, but do not confirm that of Möglich-Rompe.

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

HIGH SPEED PHOTOGRAPHY BY MEANS OF THE IMAGE CONVERTER

by J. A. JENKINS *) and R. A. CHIPPENDALE *).

778.37:621.383.8

When Holst and his co-workers invented the image converter, the fascinating thing about this device was the possibility of evading Stokes' law in fluorescence and, therefore, transforming infrared images into visible ones. In recent years, this application of the image converter has grown rather commonplace and new possibilities have been discovered, one of the latest and most spectacular being the application of the image converter as an extremely fast acting photographic shutter, allowing exposures of less than 10^{-7} second.

Introduction

The basic idea of the image converter as illustrated in *fig. 1*, is the transformation of an optical image into an image formed by electron rays, this then being reconverted into an optical image¹⁾. The first transformation is effected by means of a semi-transparent photocathode, the second by means of a fluorescent screen. The electrons emitted from the photocathode are focused to an image on the screen by a suitable electron-optical system, in which the screen is part of the anode.

The device has found large-scale application for rendering infra-red images visible to the eye²⁾. In this case use is made of the fact that light emitted by the fluorescent screen can have shorter wavelengths than the radiation forming the original image. Another important application of the device is the X-ray image intensifier³⁾. This is based on the possibility of imparting energy to the electrons from an accelerating electric field, thus intensifying the image, whilst the brightness on the screen can at the same time be increased by using a demagnifying electron-optical system.

The application of the image converter as a fast acting photographic shutter, that will be dealt with in this article, is due to still another of its inherent properties: the electron rays constituting the intermediate image can easily be controlled, i.e. suppressed and released or deflected, in which control-

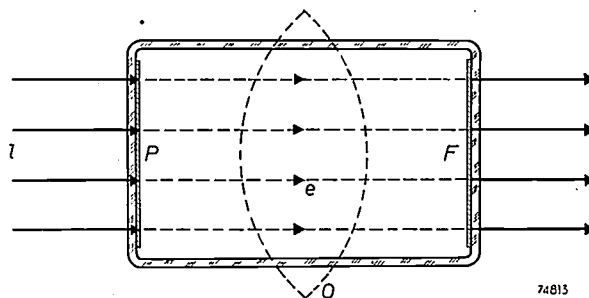


Fig. 1. Basic design of an image converter. *P* photocathode, *F* fluorescent screen, in evacuated vessel; *l* light rays, *e* electron rays; *O* symbolizes an electronoptical system.

ling action practically no inertia is involved. With the aid of an image converter tube specially developed for the purpose of high speed photography in the Mullard Research Laboratories, the ME1201 (*fig. 2*), single shots with exposure times of about 3×10^{-8} sec have been taken.

For a comparison it should be mentioned that electronic flash tubes deliver flashes of light of duration between appr. 2×10^{-4} and 2×10^{-6} sec ($2 \mu\text{sec}$). With a flash duration of $2 \mu\text{sec}$, obtained with the LSD2, the fastest of the Mullard range of

*) Mullard Research Laboratories, Salfords, Surrey, England.

1) G. Holst, J. H. de Boer, M. C. Teves and C. F. Veene-mans, An apparatus for the transformation of light of long wavelength into light of short wavelength, *Physica (The Hague)* 1, 297-305, 1934.

2) See for instance: T. H. Pratt, *Electronic Eng.* 20, 274, 314, 1948.

3) See for instance: M. C. Teves and T. Tol, Electronic intensification of fluoroscopic images, *Philips tech. Rev.* 14, 33-43, Aug. 1952.

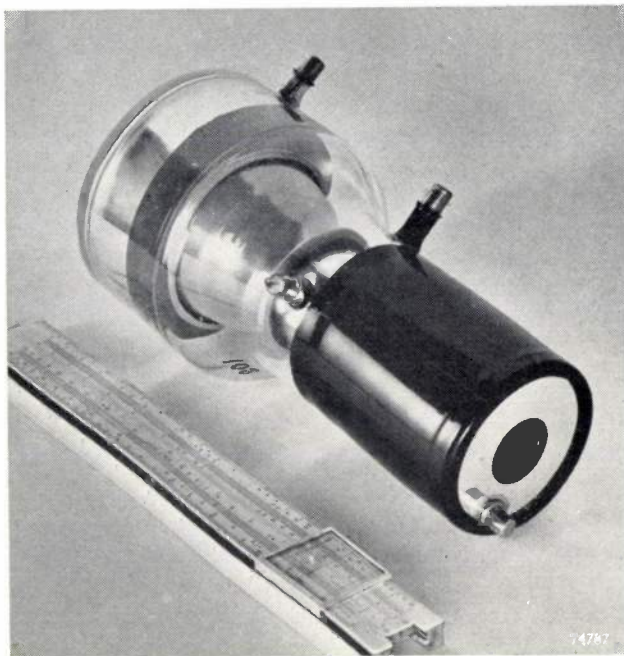


Fig. 2. The Mullard ME1201 image converter tube for pulsed operation.

photoflash tubes⁴⁾, the most extravagant requirements of amateur and professional photography in stopping fast action are by far exceeded. In the field of research, however, many instances of very rapid events are encountered which require an exposure time of this order or even less in order to obtain sharp pictures. Some examples are: the behaviour of projectiles impinging on obstacles, the vibrations and mechanical failure of rotors at high rates of revolution, fuel injection from high pressure jets, the distribution of stresses in dynamically loaded models, etc. In several of these cases, an exposure of 2 μ sec will already be far too long. In other instances, as in the study of combustion, spark and discharge phenomena, detonation of explosives, etc. the objects of the investigation are self-luminous. In these instances an illuminating flash in any case cannot be the answer to the photographic problem, but an extremely rapid shutter is needed to permit the use of exposure times that are short as compared with the light emission period of the event under examination. The exposure must then be controlled by means of a synchronization method similar in principle to the well known synchronization used in conjunction with normal flash bulbs and normal photographic shutters, but different in that extremely short delay times and triggering times are required. Special synchronizing circuits making use of hydrogen thyatrons have been developed to this end.

⁴⁾ G. Knott, High-intensity flash-tubes, *Photographic J.* **89B**, 46-50, 1949.

A brief account of the development of the ME1201 image converter mentioned above and of the synchronizing circuits will be given in this article⁵⁾.

Set-up

The set-up for applying the image converter as a photographic shutter is shown schematically in fig. 3. A sharp and bright image of the event or object to be photographed is produced on the photocathode of the image converter tube by a large aperture photographic lens of a type used in modern miniature cameras. At the other side of the tube a camera suited for reproduction work is permanently focused on the fluorescent screen. Fig. 3 also shows the triggering and synchronizing circuits that are connected to the image converter tube in order to let the electronic image pass only during the prescribed exposure time and to synchronize the exposure with the event under examination. In cases where the object is not self-luminous but is lighted for instance by an electronic flash tube, the circuits must also trigger this light source at the right moment.

The main part of this article will be concerned with the taking of single shots only. The problems involved in taking a number of pictures in rapid succession will be briefly dealt with in the last section of this article.

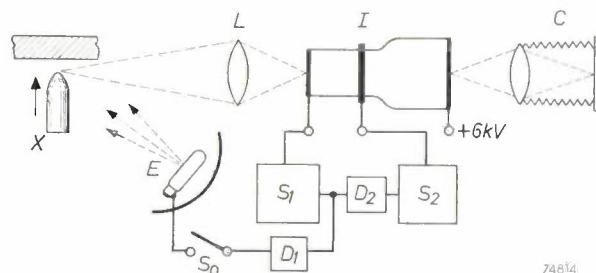


Fig. 3. Set-up for using the image converter as a high speed photographic shutter. *X* event to be photographed, *E* illuminating light source, *L* lens, *I* image converter tube, *C* photographic camera permanently focused onto the screen of *I*. The light source and the image converter are triggered and controlled in the correct intervals by means of switching and delay circuits S_0 , S_1 , S_2 and D_1 , D_2 .

⁵⁾ A first and rather extensive account was given by the same authors in *J. Brit. Inst. R.E.* **11**, 505-517, 1951 (Nov.). Cf. also: *Electronic Eng.* **24**, 302-307, 1952 (July) and *Photographic J.* **92B**, 149-157, 1952 (Sept.-Oct.). The circuits to be used in conjunction with the image converter were extensively dealt with by M. S. Richards, An industrial instrument for the observation of very-high-speed phenomena, *Proc. Inst. El. Eng.* **99** part III A, 729-746, 1952 (No. 20). — It may be mentioned that the idea of using the image converter as a high speed photographic shutter seems to have been put forth first by R. T. Bayne in a patent application filed as early as 1943 (U.S. Patent No. 2, 421, 182, May 27, 1947). Several workers independently developed the same idea, one of the foremost being J. S. Courtney-Pratt (*Research* **2**, 287, 1949).

Fig. 4 may serve to illustrate the quality of the picture obtained in a single shot of exposure time 2×10^{-7} sec.

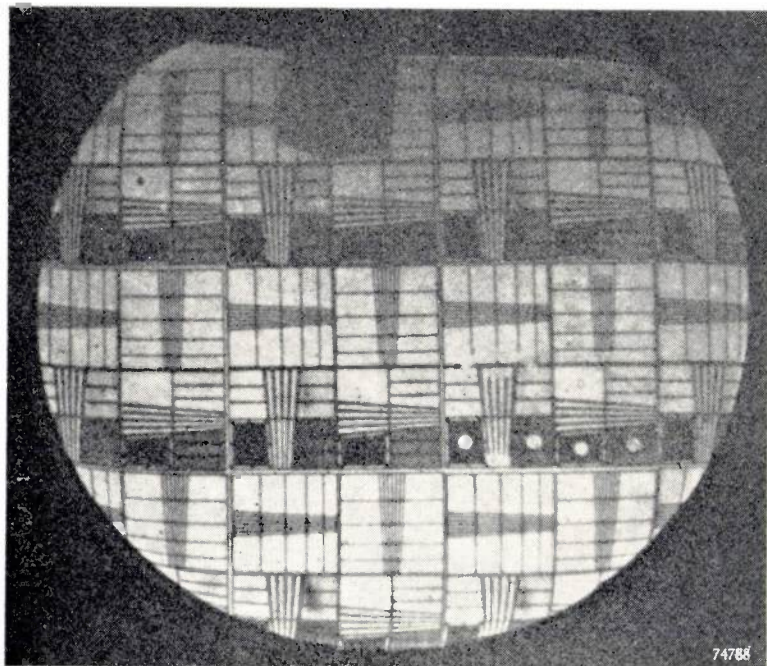


Fig. 4. Picture of a test chart recorded from the fluorescent screen of the image converter tube ME1201 with an exposure time of 2×10^{-7} sec.

Design of the image converter tube ME 1201

Grid electrode

In designing an image converter for the purpose described, the avoidance of inertia effects and the furtherance of controllability stood foremost. In the first experiments, carried out with existing standard image converter tubes, these were controlled by pulsing the high anode voltage (usually between 6 and 12 kV). Easier and faster operation is made possible by introducing an extra electrode acting as a grid with a good control characteristic. This has been done in the ME1201 (fig. 5). The tube

current is biased off by giving the grid electrode a potential of about 60 V below cathode potential, and the tube may be triggered by a high positive voltage pulse, of say 2 or 3 kV, on the grid. During operation the screen (anode) is at a potential of 6 kV with respect to the cathode.

A high grid potential is essential for obtaining at the photocathode surface an electric field intensity ensuring the maximum possible electron emission (elimination of influence of space charge). At the same time, this electric field combined with a constant magnetic field produced by a focusing coil as illustrated in fig. 6 constitutes the electron-optical system focusing the electrons onto the fluorescent screen. This system, which is characterized by low distortion and good definition of the picture obtained, is similar to that used in the image-icoscope, a type of television pick-up tube recently described in this Review⁶⁾.

The electron-optical system will be considered in more detail presently, but first a few words should be said regarding the photocathode and the fluorescent screen.

Photocathode

The semi-transparent photocathode is of the caesium-antimony type also employed in the image icoscope and in other devices, e.g. in photomultipliers. It is noteworthy that photocathodes of this type, when made according to the conventional methods, are found to be

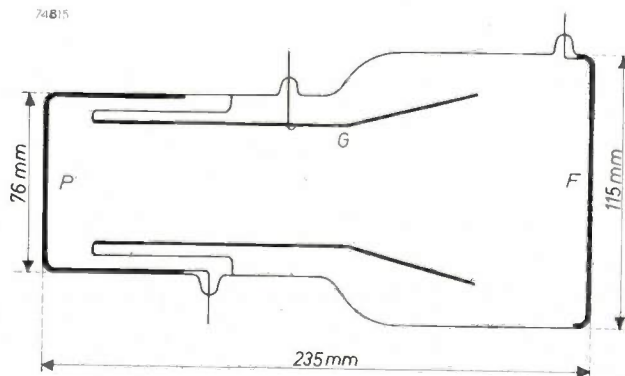


Fig. 5. Cross-section of ME1201 tube. *P* photocathode, *F* fluorescent screen, *G* grid electrode. The latter is formed by a layer of aluminium on a glass tube supported by a reentrant part of the vessel. Near *P* and *F*, the inner wall of the vessel is metallized to connect the photocathode and the metal-backed screen to the cathode and anode lead-in respectively. Part of the vessel contains an outer metallization connected to the cathode in order to minimize the occurrence of internal wall charges that will result in the transmission by the tube of spurious signals.

⁶⁾ P. Schagen, H. Bruining and J. C. Francken, The image icoscope, a camera tube for television, Philips tech. Rev. 13, 119-133, Nov. 1951.

completely unsuitable for our purpose: when it is tried with such a cathode to photograph the discharge of a fast electronic flash tube as

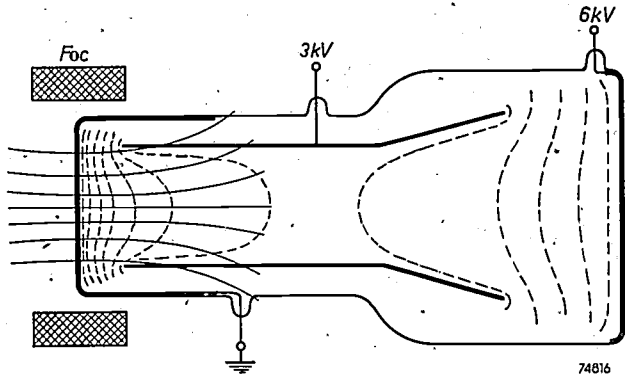


Fig. 6. Electron-optical system for the ME1201. The focusing coil *Foc* produces a magnetic field whose lines of force are indicated by thin lines. The grid and anode potentials establish an electric field whose equipotential surfaces are indicated by dotted lines. The joint action of these two fields on the electrons emitted from the photocathode produces an image of the photocathode on the fluorescent screen.

e.g. the LSD2, no picture at all becomes visible on the fluorescent screen of the image converter. This is due to the photocathode not being able to conduct in such a short interval the required number of electrons to the parts which are subjected to an intense illumination and from which the electrons should be emitted (fig. 7). In fact, the surface resistance responsible for this failure amounts to several megohms per square⁷⁾ in conventional photocathodes. A special caesium-antimony cathode was developed, having a surface resistance of

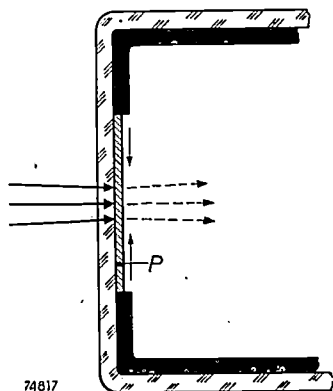


Fig. 7. The useful part of the photocathode *P* must be very thin (semi-transparent) in order to transmit the light to the inner surface, where the electrons must leave the metal (dotted arrows). In supplementing these electrons from the surrounding ring-shaped metal layer, the photocathode must carry a current parallel to its surface (thin arrows). For adequate negative densities produced with an exposure time of 10^{-7} sec, peak currents of tens of milliamperes must be drawn from the cathode.

⁷⁾ The surface resistance of a layer of given resistivity and thickness only depends on the configuration of the conducting surface, not on the dimensions.

only a few hundreds of ohms per square. This value has proved to be small enough, as now the actual limitation in decreasing exposure times in most cases is no longer imposed by the cathode. In reducing the surface resistance, part of the overall sensitivity of this type of photocathode had to be sacrificed; it amounts to $20 \mu\text{A}/\text{lumen}$ for incandescent light (from tungsten at 2700°K), as against values of 45 or even $60 \mu\text{A}/\text{lumen}$ that have been obtained with conventional photocathodes.

The spectral response curve of the low-resistance caesium-antimony photocathode is given in fig. 8. It resembles the spectral response curve of the eye, except for a shift to the blue. This response curve is ideally suited for photographing objects illuminated by an LSD2 flash tube, as may be seen from the spectral energy distribution curve of the latter

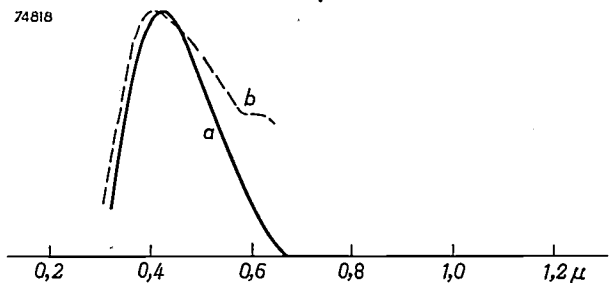


Fig. 8. a) Spectral response curve of low-resistance caesium-antimony photocathode (on a relative scale.) b) Spectral energy distribution curve of LSD2 electronic flash tube (on a relative scale).

(*b* in fig. 8). For other light sources other photocathodes can be used, for instance the well known Ag-O-Cs cathode which is sensitive to infra-red radiation, or a combination of one of these cathodes with a phosphor sensitive to ultra-violet radiation or to X-rays.

Fluorescent screen

The fluorescent screen material in most cases adopted for the ME1201 is zinc sulphide. Alternatively willemite (Zn_2SiO_4) may be used. The former radiates about 3 times more energy per second than the latter, when excited by 6 kV electrons. Moreover, the spectral distribution curve of the radiation emitted by a zinc sulphide screen is better adapted to the response curve of the high sensitivity photographic emulsions that must be used for this work (cf. fig. 9). Both effects combine to make zinc sulphide 5 times more efficient photographically than willemite, so that the former material is preferable for general purpose use. Another difference between the two materials is the slower decay of the luminescence usually exhibited by a willemite screen.

Curiously enough, this might be considered as an advantage for single shot photography. No blurring is introduced by the fact that light is being emitted from the screen for a long time after the electrons conveying the image have been cut off, and in fact this light is beneficial to the density of the negative obtained, as the sensitivity of any emulsion will decrease at very short exposure times (reciprocity failure or Schwarzschild effect). For multiple shots with a high repetition rate, however, a short afterglow duration is essential in order to avoid that each frame should suffer interference from the preceding ones. With a zinc sulphide phosphor containing nickel as a "killing" agent, decay times as short as 10 μ sec have been obtained (drop from peak intensity to 25%).

The efficiency of the fluorescent screen is greatly increased by an aluminium backing, formed in the usual way⁸⁾ and reflecting most part of the light that is emitted backwards and would otherwise be lost. At the same time, this portion of the emitted light is prevented from hitting the photocathode and producing extra electrons that would cause severe fogging of the fluorescent image. In our specific case the metal backing layer fulfils two more functions of equal importance. Firstly the caesium vapour which is present in the tube during the formation of the photocathode is prevented from contaminating the phosphor. Secondly the light transmitted by the photocathode — and which in this application of the image converter in general has about the same wavelengths as the light emitted

thin enough for transmitting practically all the 6 kV-electrons. Considerable care is needed to produce such aluminium layers not containing sections too transparent to light.

Both willemite and zinc sulphide can be made to form very fine grained screens, capable of resolving 50 lines/mm⁹⁾. This resolution is high enough to ensure that in most cases the performance of the image converter tube will not depend on the resolving power of the fluorescent screen. The following paragraphs will make this clear.

Electron-optical system

It was already mentioned that the electron-optical system comprises a magnetic field produced by a coil around the cathode part of the tube and an electric field produced by the grid and anode potentials. In fact, the electric field existing between the grid and the anode should be considered as a separate, electrostatic lens. For the moment, however, this can be neglected. The electron-optical system produces at the fluorescent screen an image of the photocathode with a magnification of about 4 diameters and with a rotation of about 20°. Such a rotation, inherent in all magnetic electron lenses, is of no consequence whatsoever in our case. An essential property of the combined magnetic and electric electron-optical system is that a nearly flat image is obtained from a flat photocathode, thus facilitating the manufacture of the tube and enabling the use of normal photographic camera optics at both ends in the set-up.

The distortion of the picture is found to depend greatly on the absolute values of both the magnetic and electric field intensities. Among different types, pincushion distortion and S-bend distortion are most prominent, the latter being typical of all magnetic electron-optical systems (cf. ⁶⁾). Both effects are proportional to d^3 , d being the diameter of the useful part of the photocathode. The important influence of the electric field intensity at the cathode is demonstrated by the series of pictures in *fig. 10*, taken with different values of the grid potential V_g . It is seen that with the cathode diameter $d \approx 2.5$ cm used in the ME1201 the S-bend and other distortions become negligible at $V_g >$ about 2.5 kV.

It should be noted that for each value of V_g (and V_a if this were changed), the focus coil current must be adjusted anew so as to bring the image in sharp focus on the fluorescent screen. It is interesting to

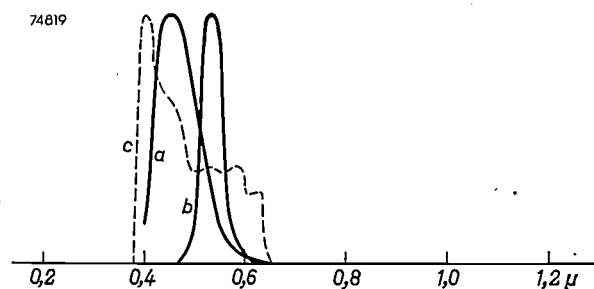


Fig. 9. a) Spectral energy distribution of fluorescence of a typical zinc sulphide screen.
b) The same for a willemite screen.
c) Spectral sensitivity of a fast panchromatic emulsion.

from the screen! — is not allowed to illuminate the screen. Evidently, this is a *conditio sine qua non*, as such an illumination would be liable completely to spoil the pictures to be taken. Even with the metal backing the avoidance of this effect is by no means easy, as the metal layer must be

⁸⁾ See for instance: Philips tech. Rev. 10, 101-102, 1948.

⁹⁾ I.e. a pattern containing 25 black and 25 white lines per mm. It is usual in television technique to designate this as 50 lines/mm, whereas according to conventional photographic terminology this would be called 25 lines/mm.

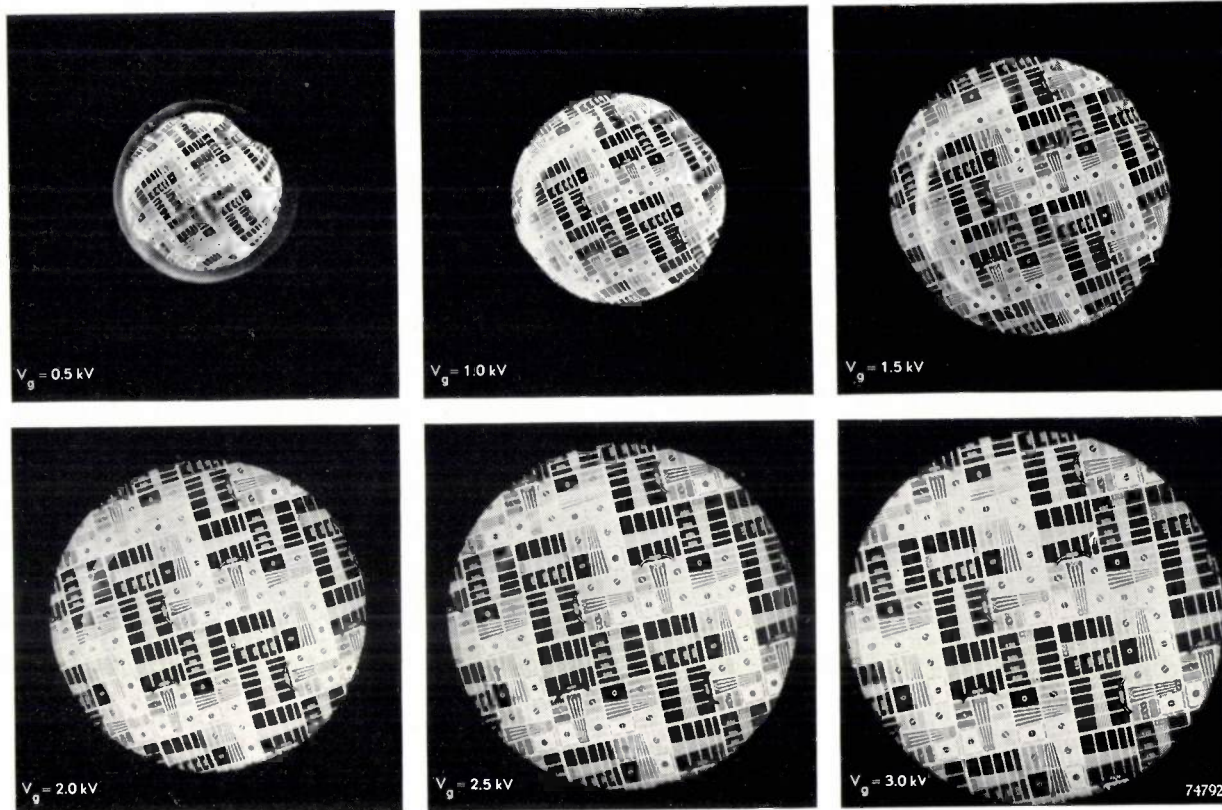


Fig. 10. Static pictures recorded from the screen of the ME1201, at an anode voltage of 6 kV and different grid voltages V_g . At low grid voltages the image reveals S-bend and other distortions.

note that for each V_g value this can be effected with several distinct current values, the successive images increasing in size with increasing current. The image obtained with the lowest current value ("first focus") suffers from considerable distortion. The pictures in fig. 10 were obtained with the second or third focus, which is also normally used and for which the above-mentioned values of magnification and rotation are valid.

The appearance of several foci can be understood from the mathematical treatment of the image formation in an image iconoscope, given by Francken and Dorrestein¹⁰). Their results can be summarized for our purposes as follows. The paths of all the electrons emitted from one point of the photocathode with different radial and/or tangential velocities intersect again in a point at a distance z_0 from the cathode. The value z_0 is found to be the same for electron pencils emanating from all cathode points, i.e. an image of the cathode is formed at a distance z_0 . In order to compute z_0 , the electron pencil originating in the center of the photocathode, whose rays will intersect on the axis of the system, is examined. Apart from a rotation, which does not matter in this case, the path of an electron starting with an initial radial energy ϵ_r is given by the function:

$$x = 2 \sqrt{\frac{\epsilon_r}{eF}} r_2(z),$$

¹⁰) J. C. Francken and R. Dorrestein, Paraxial image formation in the "magnetic" image iconoscope, Philips Research Reports 6, 323-346, 1951 (No. 5).

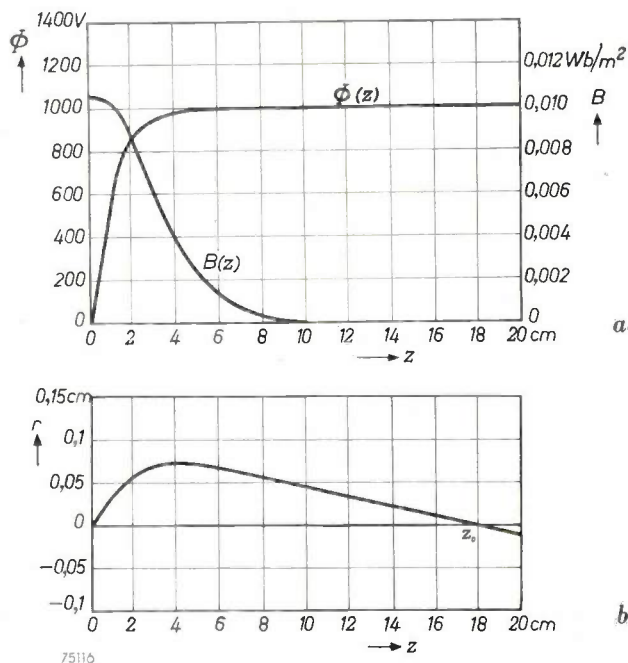


Fig. 11. a) Flux density B and electric potential ϕ on the axis (z -direction) of an image iconoscope as described in the articles quoted in ⁶) and ¹⁰). The field configuration is similar to that in the electron-optical system of the image converter ME1201.

b) Path of an electron emitted from the center of the photocathode, as calculated by Francken and Dorrestein¹⁰) for the magnetic and electric fields given in (a). The initial radial energy of this electron is $\epsilon_r = 0.65$ V; for other values of ϵ_r , the radial displacements $r(z)$ are proportional, so that all the electrons emitted from the centre of the photocathode intersect the axis at z_0 .

r being the radial displacement from the axis, F the electric field intensity at the center of the photocathode, e the elementary charge and $r_2(z)$ a function dependent on the shape of the magnetic and the electric field. The path $r(z)$, calculated for a magnetic and an electric field typical of the image iconoscope and for a practical value of the initial electron energy ϵ_r , is plotted in *fig. 11*. If the zero point z_0 of $r_2(z)$ occurs within the magnetic field region and if this field is intense enough, the rays may be turned back towards the axis again, and more zero points of $r_2(z)$ may occur, i.e. more than one image may be obtained. This evidently is our case. When increasing the magnetic field intensity, the curvature of the rays $r(z)$ is increased, all the images are drawn nearer to the photocathode and the images are successively made to coincide with the fixed plane of the fluorescent screen.

It is further shown by Francken and Dorrestein (formula 62d of their article) that the image magnification is inversely proportional to the final slope of $r_2(z)$, at the respective focus (z_0, \dots). This slope will be smaller for higher focus order, even when the focus distance in each case is made equal to the tube length. In this way the increased image size of successive foci is understood.

The definition of the electron-optical image obtained at the fluorescent screen is greatly affected by fluctuations of the magnetic and electric field intensities. Careful stabilization of the grid and anode voltages as well as of the coil current is necessary if the highest definition, matching the resolving power of the fluorescent screen, is required. *Fig. 12*, a reproduction of a static picture produced at the screen, may serve to demonstrate the optimum resolution of which the electronoptical system is capable. When microsecond photographs are being made, a lack of stability will not of course exert its influence within the very short exposure period. As the focusing, however, is effected by producing a static picture of the object prior to the actual exposure and by adjusting the anode and grid voltages and the coil current so that this picture when examined visually is sharpest, a change of the adjusted values in the time interval between the

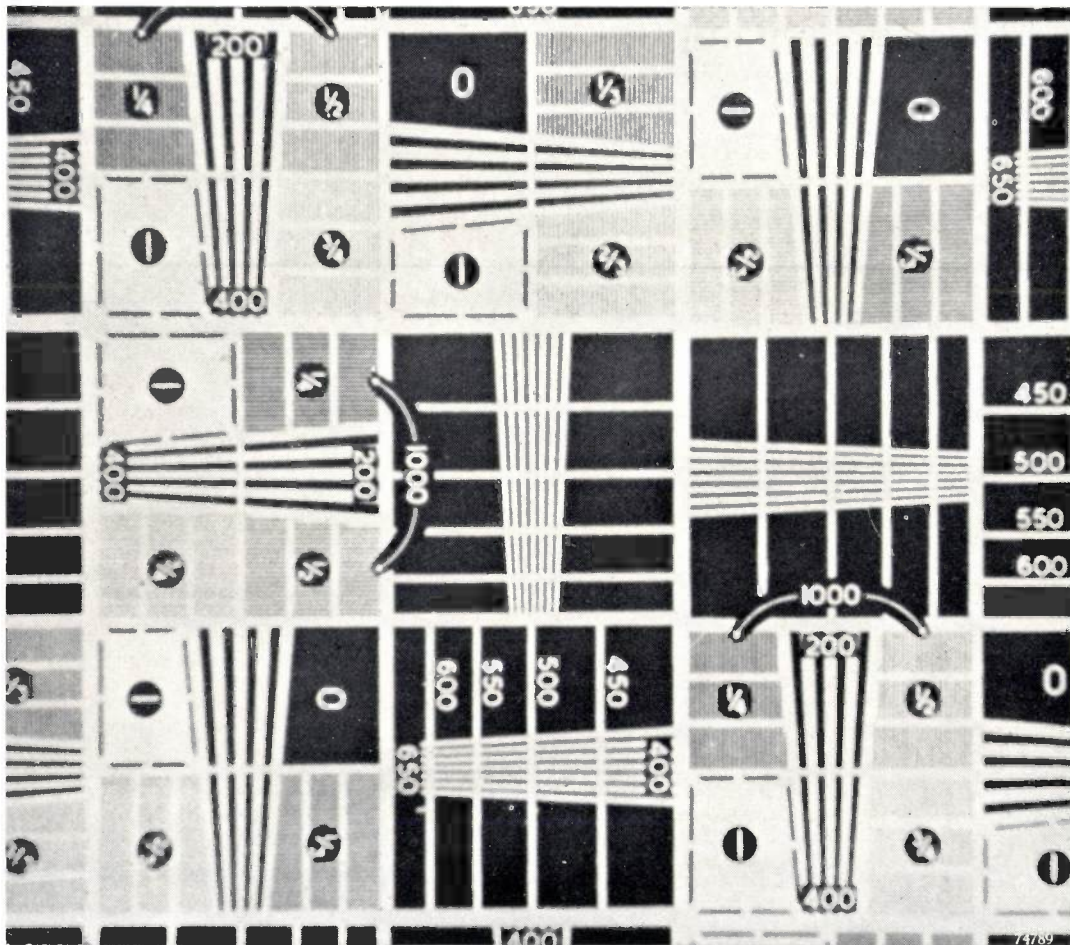


Fig. 12. Static picture of test pattern recorded from central part of the screen of the ME1201 image converter, in order to demonstrate the optimum obtainable resolution. About $3 \times$ size of image on screen. The influence of coil current and grid and anode voltage fluctuations was eliminated and the picture was recorded on a very fine-grained emulsion. The resolution is about 50 lines/mm.

adjustment, and the exposure may result in a loss of focus. The stabilization of the grid voltage is especially critical, as this will affect both the electron-optical system formed by the combined magnetic and electric fields and (to a lesser extent) the electron lens formed by the electric field between grid and anode; the focus shift caused by a grid voltage variation unfortunately is for both lenses in the same direction. However, with a ripple of about 1% in the grid and anode voltage and in the coil current, 20 lines per mm will be resolved at the fluorescent screen. For many practical purposes this will do, as it corresponds to about 2000 lines in the entire picture, with a fluorescent screen image diameter of about 10 cm. The total number of lines (the "information" conveyed) then is 4 or 5 times that of a television picture according to American or British standards.

It has been stated that the electron-optical system with the normally used focusing coil provides a linear magnification of about $4\times$. As nearly all the electrons emitted by the useful part of the photocathode are concentrated in the image at the fluorescent screen, the screen brightness will be proportional to the inverse square of the magnification. Therefore, in cases where a shortage of light is experienced, lower magnification or even a demagnification may be desirable (cf. the article quoted in ³⁾). With the aid of other focusing coils other magnifications, varying from $7\times$ down to $2\times$, are possible. For even lower magnifications, other tubes have been designed. With one of our latest tubes, which is not available as yet for regular distribution, a demagnification by a factor $\frac{3}{4}$ is obtainable, resulting in a gain of about $16\times$ in brightness of the fluorescent image as compared with the ME1201 in normal use, all other things being equal. It should, however, be pointed out that in cases where the struggle for sufficient light does not set the scene, a certain electron-optical magnification as used in the ME1201 offers important advantages ¹¹⁾. A lower magnification will almost inevitably mean a smaller image at the fluorescent screen, as it is very difficult to increase the useful photocathode area without introducing an excessive distortion in the electron-optical reproduction, to say nothing of the difficulty of avoiding local sensitivity variations in large photocathodes. A smaller screen image size, however, is bound to bring the graininess of the screen into greater prominence, affecting the

overall resolving power of the method. This point of view is especially important in a procedure for taking multiple shots that will be described at the end of this article. The loss in resolution may partly be restored in the actual recording process. In fact, as long as the shortage of light is predominant, very sensitive emulsions must be used in recording the picture from the screen and the resolution will be limited by the film grain. If the brightness of the screen image is substantially increased, finer grained emulsions of lower sensitivity may be used.

Triggering and synchronizing circuit for single shots

The general components of the circuit used for taking single shots with the ME1201 have already been indicated in fig. 3. In the "waiting" condition of the image converter the cathode potential is made 3.1 kV, the grid potential 3 kV. No electrons can leave the photocathode in this condition. The tube is triggered (the "shutter is opened") by bringing the cathode potential down to zero by means of a switch S_1 . It is biased off again (the "shutter is closed") by bringing the grid down to a slightly negative potential, e.g. -100 V, by means of a second switch S_2 . The switches S_1 and S_2 , which must be operated in rapid succession, to define the interval of the exposure, consist of hydrogen thyratrons which can be triggered at an accurately fixed time by a comparatively low grid voltage pulse.

The grid-cathode system of the image converter tube possesses a finite capacitance, amounting to appr. 100 pF, that must be charged and discharged in the successive pulsing of cathode and grid potentials. These charging and discharging processes, involving the transport of an electric charge of about $3000 \times 100 \times 10^{-12} = 3 \times 10^{-7}$ coulomb, will take some time, depending on the maximum current the thyratron tubes are able to carry without damage to their cathodes, on the building-up time of this current and on the resistance occurring in the charging circuit. The charging and discharging periods should be made as short as possible, as during these periods electrons are passed from the photocathode to the fluorescent screen. It is true that during these periods the proper focusing conditions are not established, so that these electrons will not introduce a blurring due to the rapid change or movement of the object to be photographed. They may, however, contribute a background detracting from the picture quality.

A direct current of 30 A would be required to deliver the above-mentioned electric charge in 10^{-8} seconds. The 3C45 thyratron (of the range of Mullard

¹¹⁾ Cf. also: J. S. Courtney-Pratt, Image converter tubes and their application to high speed photography, *Photographic J.* 92B, 137-148, 1952 (Sept.-Oct.).

thyratrons) used in this application is rated at a current of 35 A. However, because of the building-up time of the maximum current and because the current intensity gradually decreases when the charging of the image converter capacitance is nearing its completion, the actual duration of the

current of the 3C45 thyatron can considerably be increased beyond the rated limit for single shots or for low repetition rates.

After an exposure has been made, the anode voltages of both thyratrons will gradually return to a high value, both tubes become non-conducting by

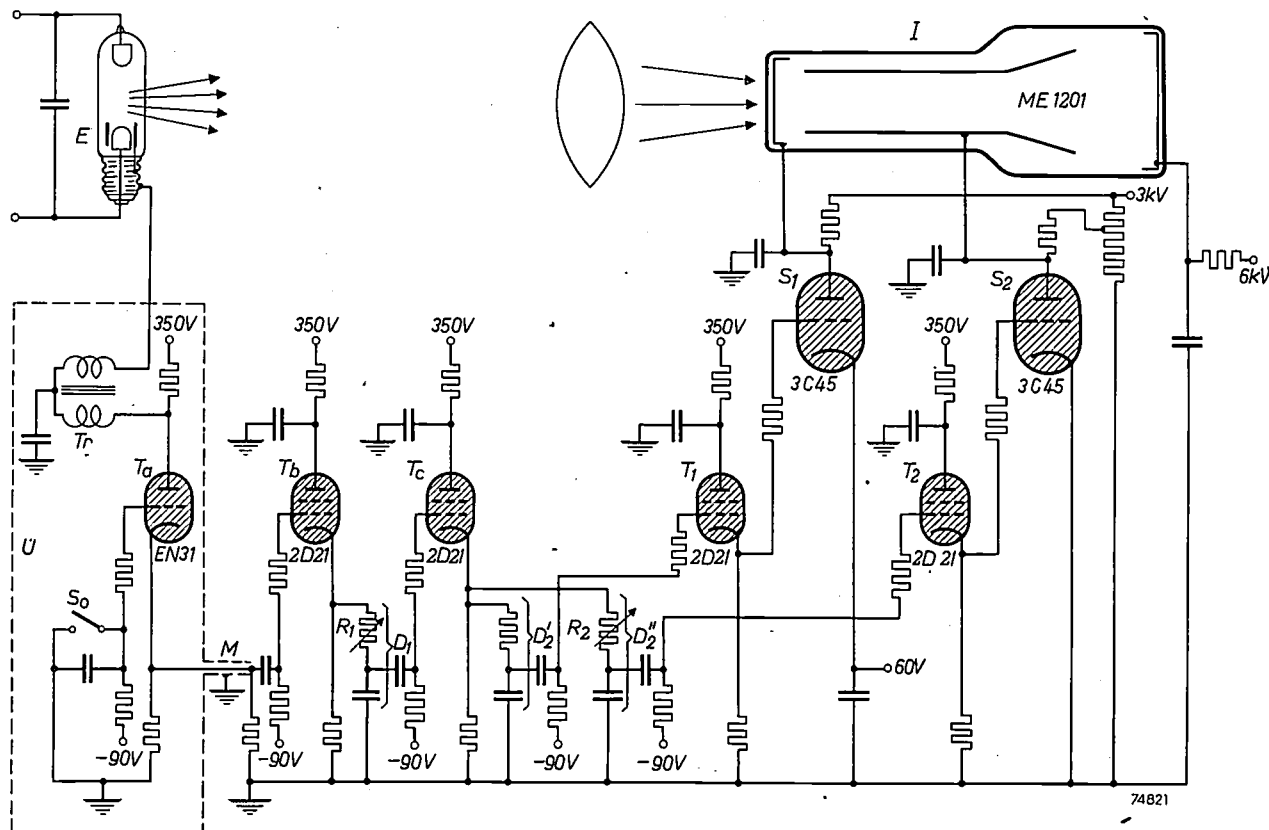


Fig. 13. Circuit for the taking of single shots with the ME1201 image converter tube. A few non-essential details have been omitted. The sequential pulsing of cathode and grid electrode of the image converter *I* is effected by the 3C45 hydrogen thyratrons S_1 and S_2 respectively (cf. also fig. 3). These are rated at 35 A and can stand a 3 kV anode voltage without breaking down at zero grid voltage. They are fired by the smaller 2D21 thyratrons T_1 and T_2 , which receive their triggering grid pulse voltages from a thyatron T_c via two delay circuits D_2' and D_2'' respectively. D_2' gives a fixed delay, whereas the delay caused by D_2'' can be finely controlled by means of the variable resistor R_2 , so that the switch S_2 pulsing the image converter grid can be actuated before or up to 2 μ sec after S_1 pulsing the image converter cathode. Thus, R_2 sets the exposure time. The thyatron T_c in turn is triggered when the mechanical switch S_0 is closed, via thyratrons T_a and T_b , and another delay circuit D_1 , adjustable by means of the variable resistor R_1 to give a delay of up to 100 μ sec. This serves to synchronize the firing of the image converter with the peak of the light emission of an electron flash tube *E*, which is triggered by S_0 via thyatron T_a and a transformer *Tr*. (The thyatron T_b is interposed because the pulse supplied by the conventional electron flash tube triggering unit *U* is not suitable for triggering T_c . The coaxial cable *M* connecting *U* to T_b must correctly terminated to avoid the occurrence of reflected signals which may fire T_b .)

charging period will be appreciably longer than 10^{-8} sec. These are, then, the limiting factors in diminishing the exposure time. The actual limit, i.e. the minimum exposure time providing pictures in which the background effect is not yet excessive, can only be determined from experience. With the present technique a minimum exposure time of 3 to 4×10^{-8} seconds has proved feasible. This result is partly due to the fact that the maximum

deionization, and the cathode and grid of the image converter are charged again to the potentials of 3.1 and 3 kV respectively of the "waiting" condition. During this recovery the image converter must constantly remain biased off; this is ensured by arranging the cathode circuit to deionize before and to charge up more quickly than the grid circuit.

The circuit developed for single shots is shown diagrammatically in fig. 13; a photograph of the

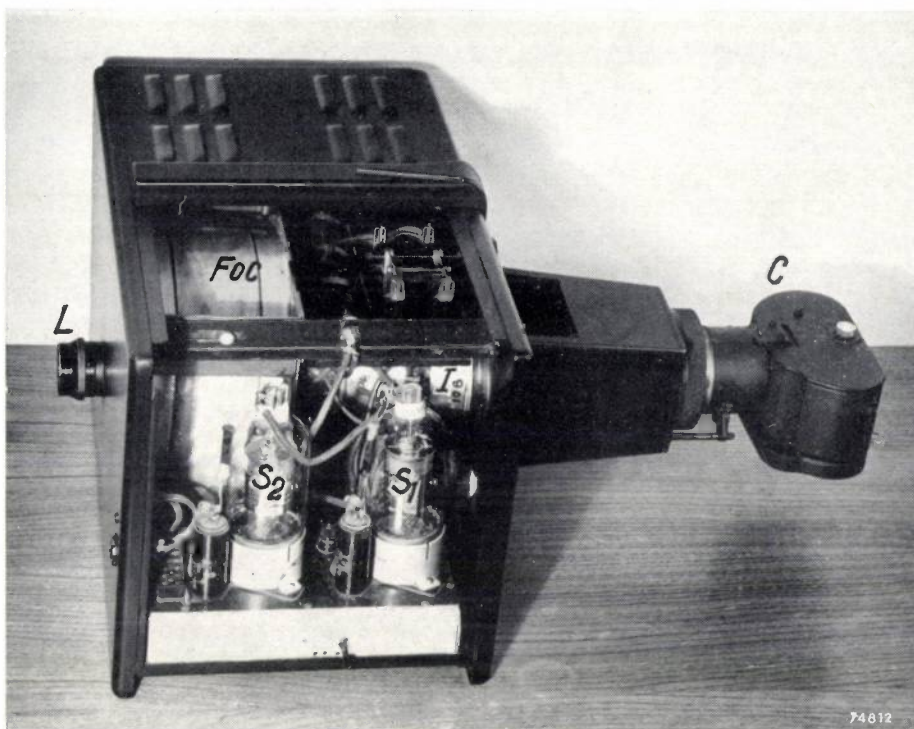


Fig. 14. Laboratory camera unit for high speed single shot photography. The image converter tube *I* is mounted in a transverse position in the cabinet. To the left the lens *L* producing an image of the subject on the photocathode is seen. The recording camera *C* projects to the right. In the near part of the cabinet the two hydrogen thyratrons *S*₁ and *S*₂ are seen.

apparatus comprising the circuit components and the image converter tube ME1201 is represented in fig. 14. From fig. 13 it is seen that the two large hydrogen thyratrons 3C45 (*S*₁ and *S*₂) are triggered by two smaller thyratrons, of the 2D21 type (*T*₁ and *T*₂). The firing of these two small thyratrons is effected by yet another 2D21 thyratron (*T*_c) via a pair of integrating circuits (*D*₂' and *D*₂''), introducing different delay times for *T*₁ and *T*₂. The time interval between the firing of *T*₁ and *T*₂ is equal to

the nominal exposure length and can be adjusted by means of the variable resistor *R*₂ to any value up to 2 μ sec. Finally, the firing of the thyratron *T*_c that starts the exposure is connected with the firing of the small thyratrons *T*_a and *T*_b in order to enable the exposure to be accurately synchronized with the peak of the luminous output of an electronic flash tube illuminating the object to be photographed. The legend of the figure gives some additional information as to this synchronization. The accuracy

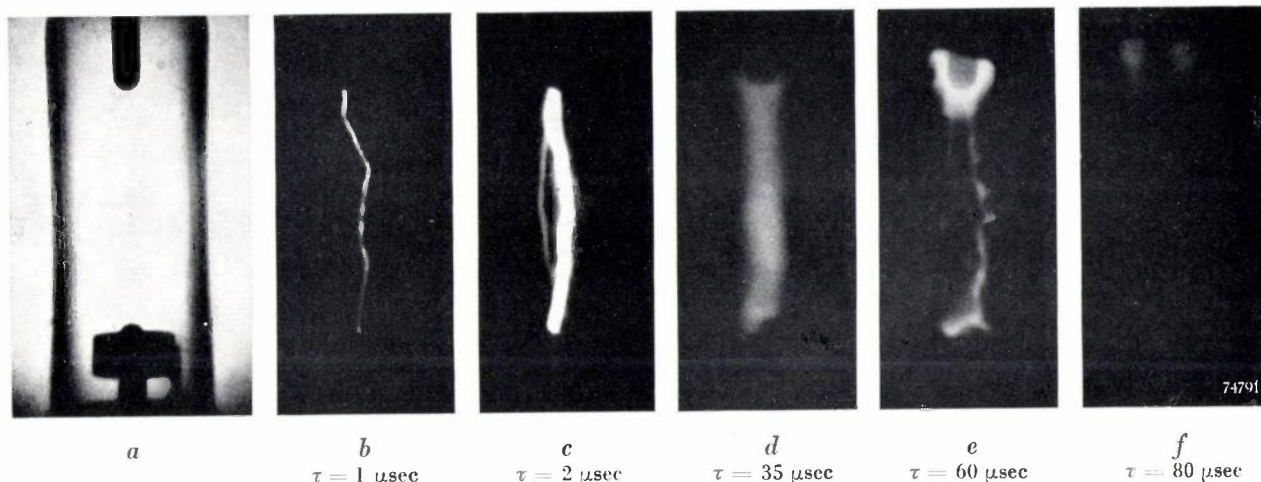


Fig. 15. Series of 0,1 μ sec exposures showing various stages in the growth and decay of the discharge in an LSD2 flash tube. *a*) is a static picture of the electrodes, *b*)-*f*) are made at increasing times τ after triggering.

of the synchronization is aptly demonstrated by *fig. 15*, showing a number of 10^{-7} sec exposures made of the discharge of the LSD2 electronic flash tube itself. The overall "jitter" is so slight that pictures of the type of *fig. 15b* can be reproduced at will in successive flashes of the tube. Incidentally, it will be clear from the pictures that this method of high speed photography is a powerful tool in the investigation of the mechanism of discharge of these electron flash tubes and the like.

Special effects in very short exposures

Several peculiarities in the behaviour of the "image converter shutter" at very short exposures have already been mentioned, e.g. the influence of the surface resistance of the photocathode delaying the supplementing of electrons to heavily emitting cathode parts; the gradual increase and decrease of the grid voltage owing to the limitations in the current supplied by the switching thyratrons; the

a relatively long time, it may in some cases be liable to ruin the record. The only general remedy for these cases will be a lowering of the anode voltage or a raising of the grid voltage.

The second effect to be described is probably again due to the surface resistance of the photocathode mentioned previously. It consists of a local distortion of the picture at places where very bright high lights occur. The effect is illustrated by *fig. 16*. Although its explanation is not yet quite certain, it seems to us very probable that it is due to the electrons emitted from the highly illuminated portions of the photocathode not being supplemented at a sufficient rate to prevent these portions from charging up to near grid potential during the very short exposure interval. Such a transient local charge may produce transverse fields at the surface of the photocathode which will exert a powerful influence on low velocity electrons emitted by other parts of the photocathode.

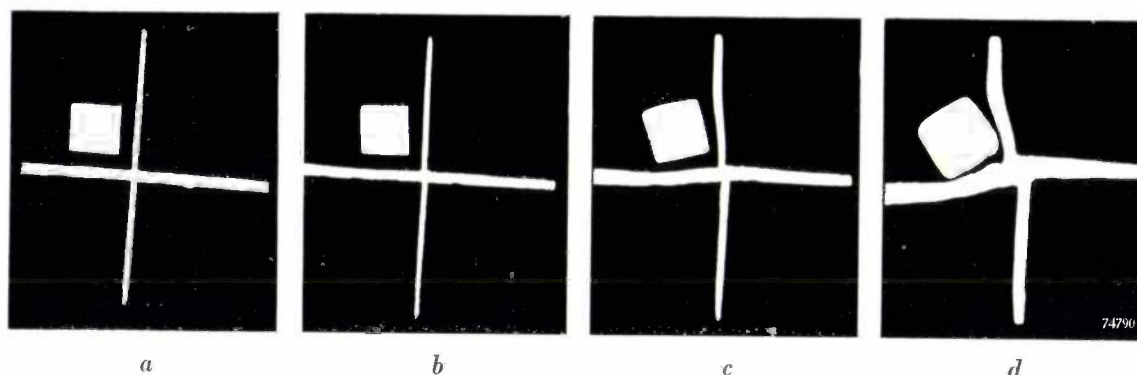


Fig. 16. Distortion due to locally overloading the photocathode of the image converter by very high illumination. A simple transparency, as shown by the static picture (*a*) was imaged onto the photocathode with different objective apertures, and exposures of $0.1 \mu\text{sec}$ were made. In (*c*) and (*d*) the illumination is $2\times$ and $4\times$ times that in (*b*), respectively.

possible transmission of light through some areas of the backing layer of the fluorescent screen during the waiting period. Two other and rather important effects should be mentioned here.

The first of these is related to the presence on the grid of minute quantities of photosensitive material deposited there during the manufacturing of the tube. If the illumination or the light emission of the object to be photographed persists during the waiting period, the light transmitted by the semi-transparent photocathode when striking the grid will liberate electrons from it. These electrons are accelerated and crudely focused by the electric field between grid and fluorescent screen, thus giving rise to a ring of light within the picture area (cf. *fig. 10*). As this effect, feeble in itself, may be integrated over

The distortion effect is rather important as it sets another limit to the minimum exposure time. From the photographic point of view this is rather strange a limit because it means that the minimum exposure depends on the brightness range rather than on the brightness of the object to be photographed. In fact, it was found in a specific case using an average image converter tube, that distortion occurred if a high light density greater than 0.5 above fog was recorded in an exposure of $0.1 \mu\text{sec}$. The practical conclusion is that for the shortest exposures the suitability of emulsions and developers should not be judged by their properties for normal use, but chiefly by their ability to bring out details at very low densities of the negative. Fortunately this ability is highest with the fast panchromatic emulsions.

Use of the image converter tube for multiple shots and scanning

If very short exposures are necessary to avoid blurring by changes or movements of the object, a cinematographic record, i.e. a continuous representation of these changes of movements, obviously will require a series of pictures with very short intervals, comparable to the exposure time. Thus frame rates between 10^6 and 10^7 per second must be aimed at. The recovery period that was mentioned in the description of the circuit used for single shots, and that is necessary inter alia for the deionization of the switching thyratrons, will prohibit any increase of the repetition rate beyond about 20 kc/s. Much higher repetition rates are made possible by the application of radar modulator techniques. In that case a limit to the shortness of the intervals between successive pictures will rather be set by the speed with which the film can be transported and by the decay of the fluorescent screen image. By the use of high-speed drums and of a suitable very fast decaying screen, film rates of say 200 000 frames/sec should be made possible. When making exposures of 10^{-7} sec, however, the intervals in this case would be still 50 times the exposure time. Thus, a real cinematographic record of the very fast events that can be stopped by the image converter method cannot be achieved in this way.

There are three different methods by which the difficulty of high frame rates can be overcome. The first one is the well-known stroboscopic method, which can be applied in the case of periodic events: successive phases of the event are not picked out from one single period but from successive periods or even hopping over a number of periods. Demonstrations of this use of the image converter have been given showing as an example the movement of stress waves in a transparent medium. A single shot circuit was used, operating the image converter shutter at 20 kc/s with $1 \mu\text{sec}$ exposures. The screen was observed visually; for this application a fast decay phosphor is not necessary, as long as the afterglow does not exceed the apperception time of the eye.

The second method consists in producing an image covering only part of the fluorescent screen, this small image being successively deflected to different positions on the screen. To this end, two pairs of deflecting coils, for horizontal and vertical deflections respectively, are fitted around the image converter tube. To each pair of deflection coils a current having one of a number of discrete intensity values can be fed, these values being alternated in such a way, that the necessary combined deflections for the different image positions are obtained. At low speeds, e.g. for deflection intervals of the order of

$1/1000$ second, a mechanical commutator can be used for alternating the deflection current values, and this will be quite a useful method for the analysis of many problems in research and industry. The same principle, however, using electronic means for alternating the deflection currents should allow of image intervals and image stopping times of the 10^{-7} sec range attained with single exposures. It will probably be possible with future types of tubes having larger screens to give the image up to 25 different positions on the screen, enabling a sequence of 25 pictures to be taken by a single shot (9 positions have been obtained already). A combination of the deflection method with the previously mentioned high speed cine-camera techniques then will give an effective multiplication of up to 25 in frame rates.

Even more powerful means for recording extremely rapid changes are available if the information desired can be conveyed by a picture in only one dimension, preferably a straight line. Such is the case with e.g. the propagation of an explosion, the starting of a discharge etc. The idea of getting separate distinct images that was adhered to in the preceding method can then be dropped and a continuous deflection of the (linear) image is possible. This continuous deflection or scanning is effected by one pair of deflection coils similar to those mentioned above. It should be mentioned that both in this method and in the preceding one, no separate pulsed grid voltage is required, and in fact a simpler converter tube (the ME1200) having no grid electrode may be used in this case. Using a sinusoidal deflection coil current, very high deflection rates can be obtained. Useful exposures were obtained by Courtney-Pratt¹¹⁾ at deflection rates of up to 300 000 metres per second; this is a writing speed about a thousand times higher than that achieved mechanically with the best drum cameras. The "time resolution" obtained with these deflection rates is of the order of 10^{-9} sec. In fact, Courtney-Pratt was able in this way to measure with reasonable accuracy the time difference in the arrival at the photocathode of two light pulses that had travelled along paths differing by a few feet in length.

Comparison with other methods of high speed photography

To conclude this article a few remarks should be made on the relative merits of the image converter technique when compared with other methods permitting exposure times of the same order¹²⁾. Among

¹²⁾ A survey has been given by J. M. Meek and R. C. Turnock, Electro-optical shutters as applied to the study of electrical discharges, Photographic J. 92B, 161-166, 9152 (Sept.-Oct.)

such methods the Kerr cell shutter is the oldest and still flourishing. When operated with electronically generated voltage pulses it permits of minimum exposure times of the same order as the image converter, i.e. 10^{-7} sec and even less, but its inherent disadvantage is the limited sensitivity. This is due partly to the light absorption in the shutter proper (the maximum transmission being only 5% of the incident light) and partly to the limited angle of the light beam transmitted. With the image converter on the contrary, the "transmission" can be well over 100% in the shutter proper, owing to the energy being imparted to the electrons by the high anode voltage, and the complete range of wide aperture, wide angle lenses can be used. In fact, the deficiency of light that was mentioned repeatedly in the preceding paragraphs is not the fault of the image converter which is a very powerful photographic device indeed, the effective aperture of the whole apparatus being higher than that of any Schmidt camera (cf. ¹¹), but it is the fault of the very shortness of the exposure times. As will be well-known to every amateur photographer, when taking an average near-by subject in bright sunshine, an exposure of the order of $1/500$ sec at $f:8$ can give a reasonable negative with a fast emulsion. Accepting a value of $f:1.4$ for the effective aperture of the set-up with image converter, and ignoring the Schwarzschild effect, the same subject brightness would require an exposure of about 2×10^{-4} sec. Thus for a 10^{-7} exposure an illumination several thousand times that in bright sunshine will be necessary.

Other methods devised for recording pictures with very short exposures make use of conventional television techniques. An image of the object to be photographed is projected onto the photocathode of a television pick-up tube, the emission of the photocathode being controlled to occur only during the desired exposure period and an optical image being displayed on a monitor cathode-ray tube. Experiments carried out by Prime and Turnock indicated that, when using a conventional iconoscope

as a pick-up tube, the suppression of the image in the waiting condition is rather incomplete and, therefore, the picture quality is spoilt by spurious image formation. Better pictures and higher sensitivity might be obtainable by using an image orthicon as proposed by Meek and Turnock ¹²), in which case the image section of the orthicon should be controlled. The device is then fundamentally similar to the image converter method described in this article, except for an increase in sensitivity obtained by the amplification section of the orthicon. It will hardly be necessary to point out that the complete set-up for this method will be much more complex and more expensive than that for the image converter method.

Summary. As the electron image in an image converter can be suppressed and released without any inertia being involved, this device can be used as a photographic "shutter" for extremely short exposures. Exposure times as short as 3×10^{-8} seconds have been obtained with a special image converter tube, the ME1201, developed in the Mullard Laboratories at Salfords (Surrey, England). This image converter tube is provided with a grid electrode, which during operation is given a potential of 3 kV with respect to the cathode, the anode voltage being 6 kV. Moreover the ME 1201 contains a special photocathode having a very low surface resistance, and a fine-grain fluorescent screen with metal backing. The electron-optical system is similar to that of the image iconoscope; it produces an image of the photocathode (useful diameter 2.5 cm) with magnification between $2 \times$ and $7 \times$ (normally about $4 \times$) and practically free from distortion. The optimum resolution obtainable in a static picture is 50 lines/mm (25 black + 25 white ones); in practical cases, with a ripple of 1% in the voltages and currents, the resolution can be 20 lines/mm. The minimum exposure time depends i.a. on the rate of charging the grid-to-cathode capacitance of the image converter, on the limited brightness of the subject to be photographed and on the brightness range of the subject; this is explained in the article. Applications of the instrument with exposure times of 10^{-7} sec have already been made in a number of investigations of objects changing or moving at a very high rate. The object may be self-luminous (explosions, sparks, discharges) or it may be illuminated by a light source such as the LSD2 electronic flash tube, having a flash duration of 2 μ sec (projectiles, high pressure jets, stress waves, etc.). A circuit for controlling the image converter tube and for synchronizing its operation with that of the light source is described in the article. For cinematographic records the possibility of giving the image a number of different positions, e.g. 25, on the fluorescent screen by means of deflection coils might be very useful. With a one-dimensional image, the method of continuous deflection (scanning) has been applied, permitting of writing speeds up to 300 000 m/sec and of a time resolution of 10^{-9} sec.

HIGH-POWER TRANSMITTING VALVES WITH THORIATED TUNGSTEN CATHODES

by E. G. DORGELLO.

621.385.33.026.455:
621.385.33.032.461:
621.385.33.032.42

Until recently the cathodes of transmitting valves delivering 100 kW or more have had to be made of pure tungsten, even though more highly emissive materials have been known for some decades. The development of very active getters, however, has made it possible to produce a transmitting valve with a thoriated tungsten cathode. Cooled with water, this valve will deliver an H.F. output of more than 100 kW with a very high efficiency. Although primarily intended for frequencies up to 30 Mc/s, the valve has also yielded good results in an experimental T.V. output stage at a frequency of 68 Mc/s.

Tungsten and thoriated tungsten as cathode materials

In the last ten years an important development has taken place in the design of large transmitting valves. Formerly, valves delivering 100 kW or more were always fitted with a filament of pure tungsten serving as a cathode. The limited emission of such a cathode necessitates a very high anode voltage; for example the TA 18/100 and TA 20/250 which, although out of date, are still in use today, operate at 18 to 20 kV on the anode. The high field strength prevailing in such valves tends to produce the troublesome phenomenon known as Rocky Point effect¹⁾, which takes the form of discharge between the electrodes at irregular intervals. These discharges not only lead to interruptions in the service, but also to untimely failure of the valve. Efforts have been made to avoid this by designing valves to operate at lower anode voltages, which in turn necessitated the use of cathodes with a higher emission than in valves for the same power with higher anode voltage. With tungsten cathodes this means that the filament consumption runs into some tens of kilowatts.

A better solution is found in the use of a cathode material having a higher specific emission than pure tungsten, for example tungsten containing about 2% of thorium oxide. Thoriated tungsten is employed for the cathodes of many small transmitting valves²⁾³⁾ and certain X-ray valves⁴⁾, but its use in large valves was formerly precluded by the high sensitivity of the cathodes to "poisoning"; the

extremely thin skin of thorium oxide responsible for the emission is attacked by the slightest traces of gas. Many attempts to incorporate this type of cathode in the larger transmitting valves, in which the vacuum cannot meet the same requirements as in smaller valves, have come to grief as a result of this poisoning effect.

The fact that thoriated tungsten has nevertheless been used with success in the valves described in the following paragraphs may be ascribed to the very high activity of the getters employed. No fewer than three different, carefully prepared getters have been introduced into the valve at suitable points, ensuring a very high vacuum and, in consequence, high and constant emission. This is amply borne out by the fact that such valves worked for years without giving any trouble, both in broadcasting transmitters and in H.F. generators for industrial use (heating).

General description of the valves

The new transmitting valves are manufactured in two types, one water-cooled (TBW 12/100) and one air-cooled (TBL 12/100). The first of these is depicted at the left in *fig. 1*; on the right will be seen an older type of valve, the TA 20/250⁵⁾. Both the TBW 12/100 and the TA 20/250 deliver a peak output of about 250 kW in telephony transmitters with anode modulation. The new valve is at once

⁵⁾ The type-numbering code is as follows: Letters. The letter T indicates that the valve is a triode (Q = tetrode, P = pentode). "A" denotes that the cathode is made of pure tungsten, "B" of thoriated tungsten (C = directly heated, E = indirectly heated oxide-coated cathode). The letter W shows that the valve is water-cooled; in the case of the TA 20/250 the W was omitted because it was obvious at the time that such a large valve would be water-cooled. Forced air-cooling is indicated by the letter L. Figures. The figures preceding the oblique stroke give the direct anode voltage in kV, and those following the stroke (very roughly) the effective output power in kW delivered by the valve continuously without modulation.

¹⁾ J. P. Heyboer, A discharge phenomenon in large transmitter valves, Philips tech. Rev. 6, 208-214, 1941.

²⁾ E. G. Dorgello, Several technical problems in the development of a new series of transmitter valves, Philips tech. Rev. 6, 253-258, 1941.

³⁾ E. G. Dorgello, Glass transmitting valves of high efficiency in the 100 Mc/s range, Philips tech. Rev. 10, 273-281, 1949.

⁴⁾ J. H. van der Tuuk, Philips tech. Rev. 3, 204 and 205, 1946.

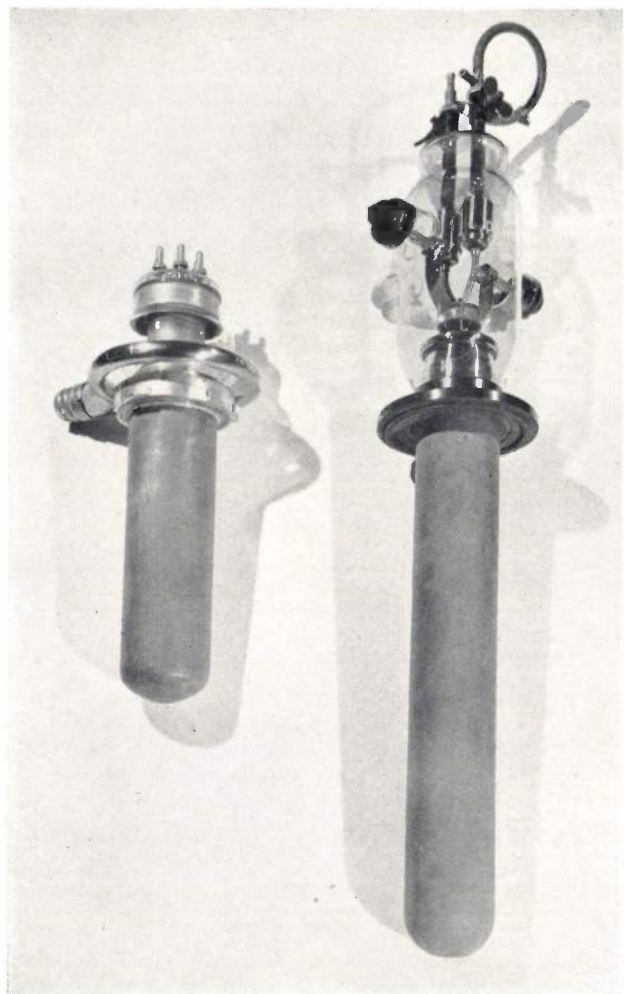


Fig. 1. Left: a modern 100 kW transmitting valve, type TBW 12/100.

Right: Older type of transmitting valve, TA 20/250. When used in a transmitter with anode modulation, they can deliver the same peak output power (250 kW).

remarkable for its smaller dimensions, from which it may be correctly assumed that the losses are smaller and the efficiency higher than in the older type. The type number of the valve reveals the lower anode voltage, viz. 12 in place of 18 or 20 kV.

Getters

Fig. 2 shows the cathode of the valve, consisting of 12 filaments of thoriated tungsten. Within the circle of filaments there is a molybdenum cylinder 3, which serves as the support for one of the three getters which takes the form of a layer of zirconium on the wall of the cylinder; radiation from the cathode heats this layer to a temperature of about 800 °C, resulting in a powerful gettering effect, at any rate as regards oxygen and nitrogen⁶⁾. The quantity of zirconium present is sufficient to

bind all the gas that the valve would be capable of holding if it were filled with N₂ at a pressure of 6 mm Hg or O₂ at 10 mm Hg, that is, one million times more than the amount of gas usually remaining after exhausting.

This getter therefore has a very high capacity but, at the same time, it works rather slowly and has to be supplemented by another which works rapidly but need not absorb such large quantities of gas.

This second getter consists of a thin layer of zirconium prepared in a different manner from the first and applied to the grid. It assumes the temperature at which it will take up nitrogen and oxygen only when the driver stage, which causes the grid to dissipate energy, is operating. The greater the amount of power supplied to the grid, the higher the temperature of the grid itself and the more rapid the absorption of gas. The maximum permissible continuous grid dissipation is 2.5 kW, and

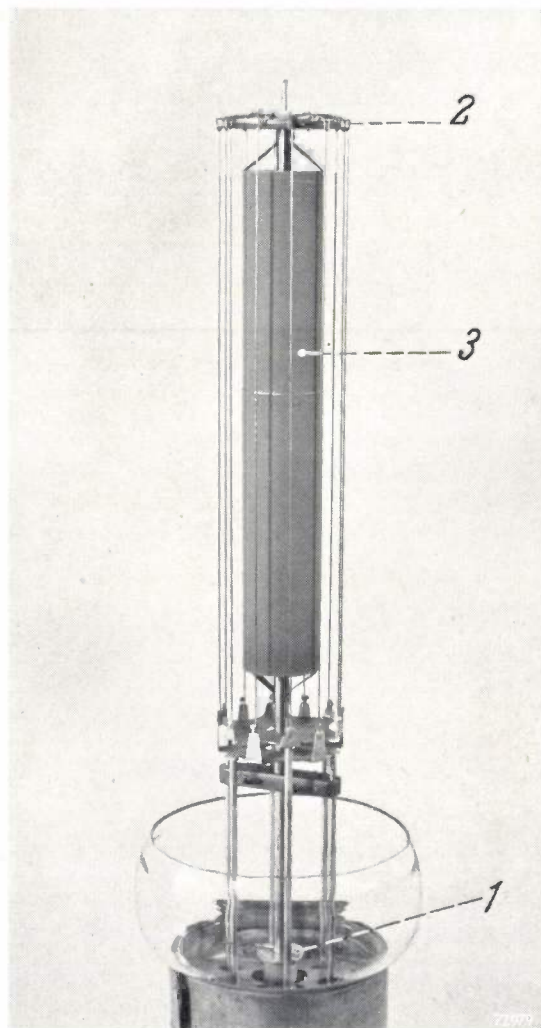


Fig. 2. The cathode of the TBW 12/100 and TBL 12/100 consists of 12 filaments of thoriated tungsten. 1 filament tension spring. 2 springs for absorbing small differences in the filament lengths. A getter 3 is mounted inside the circle of filaments.

⁶⁾ J. D. Fast, Metals as getters, Philips tech. Rev. 5, 217-221, 1940. See also the article referred to in footnote ²⁾, p. 255 onwards.

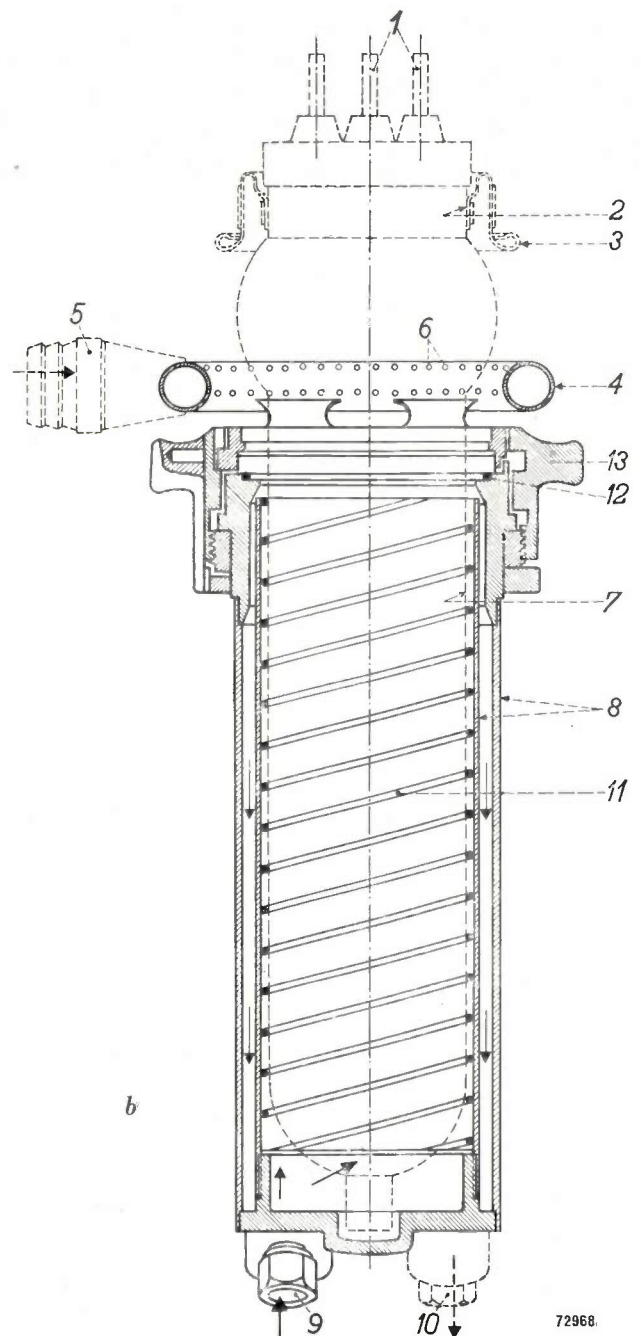
if this limit is exceeded too far, or for too long, there would be a risk of zirconium alloying with the core metal, with a consequent drop in the original high radiation factor of the zirconium layer. The temperature would then rise to a point where grid emission would be intolerably high. However, a special method of preparation of the zirconium powder used ensures that the temperature level at which alloying takes place is increased from roughly 1200 °C to about 1600 °C, which temperature occurs only with a grid dissipation of 5 kW, i.e. twice the maximum rated value.

The two getters described maintain the required high vacuum very effectively as far as nitrogen and oxygen are concerned, but at the high working temperature of these getters hydrogen is not absorbed (see articles referred to in footnote ⁶). In

order to remove any possible traces of hydrogen in the valve, a third getter is employed, again zirconium, but this time so processed and at such a working temperature that it will readily absorb this gas. This getter becomes active at temperatures below 200 °C and is accordingly placed at the bottom of the anode, in the form of a disc mounted in such a way as to be in close contact with the cooled wall of the anode. The temperature on the inside of the wall is at most 100 °C in the water-cooled valve (max. anode dissipation 100 kW), or 190 °C in the air-cooled type (max. anode dissipation 45 kW). In both cases residual hydrogen is effectively absorbed.



a



b

Fig. 3. a) The TBW 12/100 with "Grip-o-matic" water jacket. b) The same in cross-section. 1 filament connecting pins. 2 ring-shaped grid lead-in with upper anti-corona ring 3. 4 lower anti-corona ring, at the same time serving as handle and distributor of the air applied to inlet 5 for cooling the seals. 6 outlet holes for cooling air. 7 anode. 8 water jacket with water inlet 9 and outlet 10. 11 helical ribs. 12 self-sealing packing ring. 13 ring for locking the valve in the cooler.

From tests on an experimental valve, so arranged that small quantities of different gases could be introduced into it, it was found that even traces of gas may give rise to breakdown, but that the joint action of the three getters, combined with the relatively low anode potential (10 kV direct voltage with anode modulation) completely eliminates this risk.

Details of design

To ensure high and constant emission, the cathode is subjected during exhausting to the carbonising process usual with this type of cathode; this consists in burning the filament in an atmosphere of hydrocarbon to form tungsten carbide in the wire and thus improve the life of the filament. It is a well-known disadvantage of this carbonising that the wire is thus rendered rather brittle, but any risk of breakage, especially during transport, is considerably reduced by omitting carbonising at the ends of the filaments, where the greatest bending strains occur.

Deformation of the cathode and grid as well as of their supports is avoided by employing tungsten or molybdenum for all these components, both of these metals being highly rigid. The filaments are kept taut by a central spring (1 in fig. 2) made of a special tungsten alloy which retains its resilience even at elevated temperatures; moreover, this spring is mounted at a point where its temperature will remain as low as possible. Small discrepancies in the lengths of the filaments, which are unavoidable in spite of the greatest care taken in assembly, are equalized by "hairpin" springs 2.

The total filament current is 196 A, and this is applied through three lead-in pins in parallel which require no artificial cooling, in contrast with the TA 20/250 (fig. 1, top right hand; filament current 420 A), which requires water-cooling.

Connection to the grid is effected by means of a "ring-seal" (fig. 1): the self-inductance of the lead-in is accordingly low, this being an advantage when the valve is operating at very high frequencies. Owing to the large diameter of the ring, resistance to the capacitive grid current is small.

All the lead-in wires are either copper-plated or silver-plated, in order to keep losses as low as possible.

The thick-walled anode is made of O.F.H.C. (oxygen-free high conductivity) copper; in the air-cooled type the anode is provided with cooling fins, to which reference is made later.

As already mentioned, the risk of breakdown within the valve is excluded, but the fact has to be taken into account that breakdown can also

occur outside the valve, through the atmosphere, in particular between the ring-seal of the grid and the upper rim of the anode. Breakdown of this kind may destroy the glass seals and, in order to avoid this possibility, each of the components concerned is provided with an anti-corona ring (fig. 3). The electric field between these rings is such that any flashover will take place more readily between these rings, where it can do no harm, than between metal parts that are sealed to the glass. The lower anti-corona ring simultaneously serves as a handle when the valve is taken from, or inserted in the cooler, thus avoiding strain on the glass seals.

Electrical data

It has already been said above that, since the saturation current is very high, the valve is able to operate at a fairly low anode voltage, but another condition has also to be met in order to achieve this, viz. low internal resistance.

In general, low internal resistance is ensured by using a cathode having a large effective area, in which respect a large number of filaments of thin wire is better than a small number of thick filaments; however, physical strength determines the extent to which this sub-division of the cathode can be carried out.

The smallest possible spacing of anode and grid, and especially of grid and cathode, also makes for a low internal resistance, but here again considerations of a physical nature impose a limit, for, amongst other things, the valve must be able to withstand transportation over long distances.

Another consequence of increased cathode area combined with reduced inter-electrode spacing is that the capacitances are increased; this results in a drop in the maximum frequency to which the valve can be tuned, and a reduction in the band-width for a given power gain.

It follows, therefore, that a compromise must be found. Theoretically, it can be shown that low internal resistance is obtainable without entailing excessive inter-electrode capacitances by rendering the current density in the valve high. Hence a highly emissive cathode material such as thoriated tungsten is all to the good. The actual internal resistance of the valves under review is about 200 Ω .

At a filament consumption of about 3.5 kW (17.5 V \times 196 A), the saturation current of the cathode is approximately 250 A, but, to allow a safety margin, the maximum values of anode and grid current have so been specified that saturation of the cathode is not by any means reached. The maximum values will be found in *table I*.

Table I. Values of the anode current ($I_{a\max}$), grid current ($I_{g\max}$) and the corresponding anode dissipations (W_a) for the transmitting valves TBL 12/100 and TBW 12/100 at which values some quantity reaches its maximum permissible value.

Application	$I_{a\text{ peak}}$ A	$I_{g\text{ peak}}$ A	W_a kW	Cooling
Telegraphy	50	18	36	air
Anode modulation	72*)	25*)	30	air
Class B operation (linear amplifier)	80*)	30*)	95	water

*) These values, which are well on the safe side, occur only at the peaks of the deepest possible modulation.

Because of the low internal resistance, losses in the valve are also quite low. In a telephony transmitter the valves will operate at more than 80% efficiency, which is very high for a valve of this output. The value given refers to frequencies up to 30 Mc/s (wavelength 10 m); at higher frequencies the efficiency falls off gradually, although even at 68 Mc/s ⁷⁾ it is still some 60% (fig. 4).

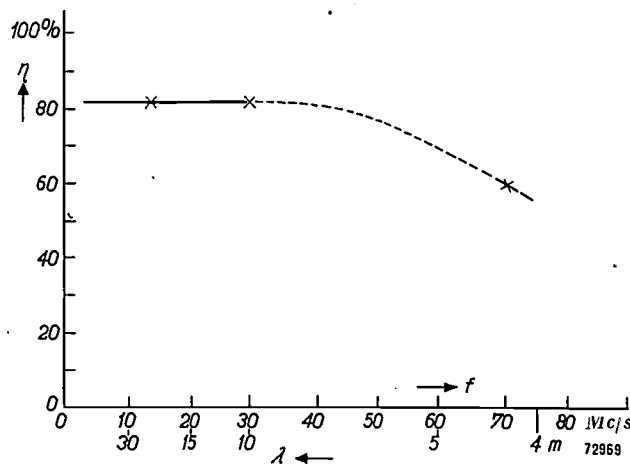


Fig. 4. Efficiency η of the TBW 12/100 operating under full load in a telephony transmitter, as a function of the frequency f or the wavelength λ . The crosses indicate the measured values.

The higher the frequency, the smaller the input power that may be applied; otherwise the displacement currents set up by the valve capacitances tend to overheat the lead-in conductors. When air-

⁷⁾ In the Philips Laboratories at Eindhoven experiments have been carried out with two TBW 12/100 valves in push-pull serving as the output stage for an experimental television transmitter. Negative modulation was employed, with frequencies up to 68 Mc/s. Output power at the sync. pulse peaks was found to be about 100 kW. The bandwidth of the anode circuit was ample for the 625-line system.

These tests will be described in a later issue of this Review. — Ed.

cooling is employed, it is a simple matter to arrange for a flow of air to pass through slots in the cooler on to the lead-in conductors of the valve.

Further electrical data relating to these valves are contained in table II.

Table II. Some electrical data of the transmitting valves TBL 12/100 and TBW 12/100.

Filament voltage	17.5 V
Filament current	196 A
Amplification factor	27 approx.
Mutual conductance (max.)	120 mA/V approx.
Capacitances:	
anode-to-grid	86 pF
grid-to-cathode	116 pF
anode-to-cathode	3.4 pF
Anode dissipation:	
with air-cooling	max. 45 kW
with water-cooling	max. 100 kW
Max. output power	
telegraphy	105 kW*)
telephony with anode modulation (carrier)	65 kW**)

*) Anode voltage 12 kV

***) Anode voltage 10 kV

For class B amplifiers an anode voltage of 15 kV max. is permissible at frequencies up to 4 Mc/s.

Cooling

Water cooling

The cooling jacket of a transmitting valve should be capable of conveying away the dissipated energy effectively and economically. In this respect, however, jackets have in the past left a great deal to be desired; with the usual construction water flow was such that only a small amount of turbulence occurred. As a result, only a thin layer of water came into actual contact with the anode surface. In order to keep the temperature of the water below 90 °C (at higher temperatures scale is formed), large quantities of water were needed, viz. about 4 litres/minute per kW dissipated.

A great improvement in economy is effected by arranging for the greatest possible turbulence of the water; each particle of the water then comes in contact with the anode again and again. A description was given in a previous issue of this Review of a jacket in which turbulence is increased by spraying the water against the anode in jets ⁸⁾, but in the jacket used with the TBW 12/100 the required result is obtained more effectively still. Owing to the helical ribs *II* on the inside of the jacket

⁸⁾ M. J. Snijders, A transmitting valve cooler with increased turbulence of the cooling water, Philips tech. Rev. 10, 239-246, 1949.

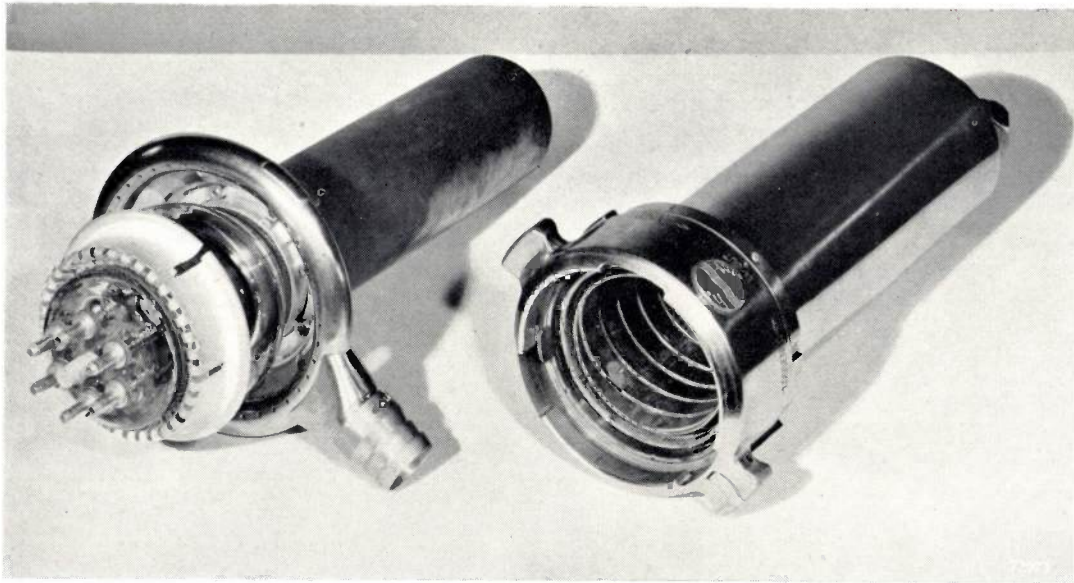


Fig. 5. Left: The TBW 12/100. Right: the water jacket in which the helical ribs — *II* in fig. 3*b* — are clearly visible.

(fig. 3*b* and fig. 5), in conjunction with the oblique positioning of the inlet and outlet, the stream of water is primarily helical in shape around the anode, but it also flows between the anode and the ribs, which then function as obstacles, thus ensuring a strong and uniform turbulence. With this jacket less than 1 litre/min per kW is required.

From the practical point of view it is important that the valve can be quickly replaced. With earlier cooling jackets this usually took a considerable amount of time; the maintenance engineer had to loosen a number of bolts with a spanner before the valve could be replaced, and this could not always be done by one man alone. The bolts had then to be tightened again to prevent leakage.

The process is very much simpler in the case of the TBW 12/100; to loosen the valve from the jacket it suffices to turn a ring (13, fig. 3*b*) through a small angle by hand; the valve is then tilted out of the jacket and the new one inserted, this being locked in position by turning the ring back. When the valve is inserted, the flange of the anode lies against a packing ring of O-shaped cross-section, and the pressure of the water in the jacket deforms this ring in such a way as to ensure a perfectly tight joint.

For further details of this system, known as the "Grip-o-matic" system, which necessitates no tools whatsoever, we would refer to another publication⁹⁾

Because of the shorter length and lighter weight of the valve (see table III), replacement of the



66169

Fig. 6. The TBL 12/100 showing the innumerable fine cooling fins on the anode ("Multifin" system).

⁹⁾ A. G. Roberer and W. L. Vervest, The "Grip-o-Matic" water jacket for large water-cooled transmitting valves, *Communication News* 12, 10-14 1951 (No. 1).

Table III. Length and net weight of transmitting valves TA 20/250 and TBW 12/100.

	Total length (with cooling jacket) cm	Length of anode cm	Weight of valve kg
TA 20/250	139	70	22
TBW 12/100	71	39	14

TBW 12/100 is very much easier than that of the TA 20/250 and the chances of accidents during replacements are so much the smaller.

The lower anti-corona ring is made in the form of a tube (see 4, 5 and 6, fig. 3b) and, at frequencies beyond 6 Mc/s, air is blown into this ring to provide extra cooling; the air streams through a number of holes in the ring and cools the glass-to-metal seals.

Air cooling

The cooler house of the TBL 12/100 is designed

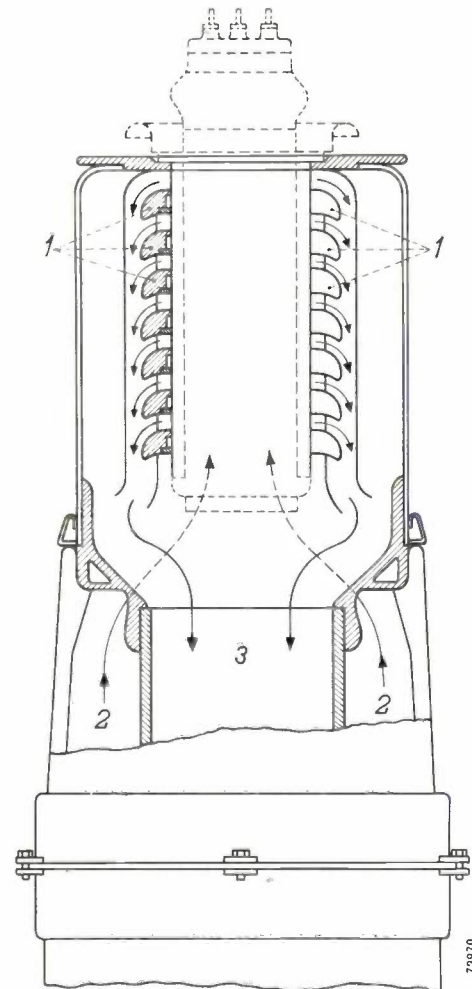
on the following principle¹⁰). It incorporates a very large number of longitudinal cooling fins on the anode surface (fig. 6), whence the name "Multifin" system. These fins are very thin and only 1 cm high; they are manufactured on a specially designed folding machine and are made from thin sheet copper, being subsequently silver-soldered to the anode. The resultant cooling area is much greater than that of the heavier fins used in other cooling systems.

A more fundamental difference, however, is to be found in the fact that the air streams in a different direction from that followed in other systems, in which one flow is used, streaming upwards along the cooling fins. Particularly when the anodes are long, the top part usually receives "cooling" air which has already attained quite a high tempera-

¹⁰) For further details see H. de Brey and H. Rinia, An improved method for the air-cooling of transmitting valves, Philips tech. Rev. 9, 171-178, 1947.



a



b

Fig. 7. a) The TBL 12/100 in its cooler. b) The same in cross-section. The baffles 1 divert the incoming air 2 into a number of short paths which come together in the outlet tube (3).

ture in the lower portions, and the cooling of the upper region is thus not very effective. In the "Multifin" system the incoming air is split up into a number of parallel paths (fig. 7), each of which covers only a small zone of the anode, so that each section receives fresh, not pre-heated air.

Here again, care has been taken to secure considerable turbulence; each jet of air is directed radially towards the anode, is then deflected upwards and downwards, and is finally led away radially; the two changes of direction ensure ample turbulence.

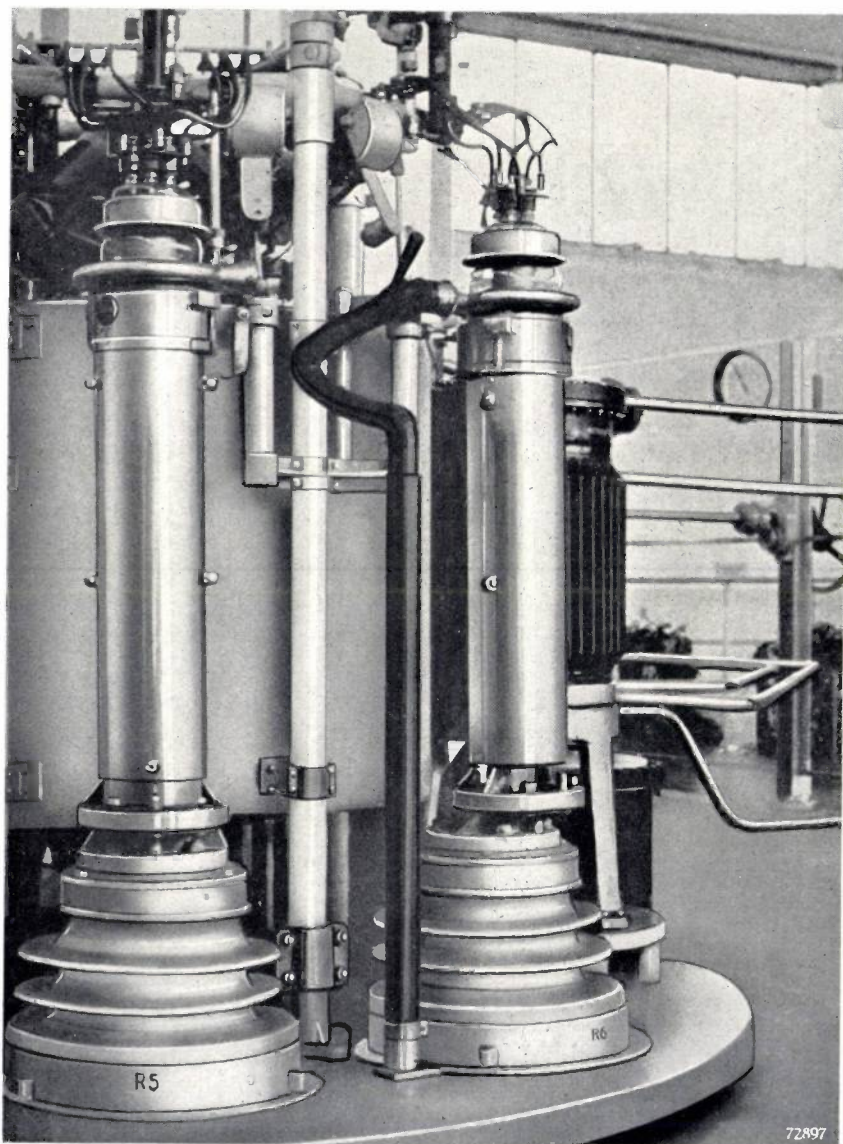


Fig. 8. In the foreground (right) a TBW 12/100 in the 100 kW broadcast transmitter at Mühlacker near Stuttgart.

By this method 50 W to 60 W per cm^2 of anode area can be dissipated for a consumption of about $1 \text{ m}^3/\text{min}$ of air per kW, without the temperature of the inner anode wall rising above 180°C . Compared with the earlier system, this is an improve-

ment in the air consumption by a factor of 2, at the same anode temperature. The used air is led out through a central tube (3, fig. 7b). This has two advantages in comparison with the older system of merely allowing the air to escape; in the first place the atmosphere in which the valve is mounted is not heated unnecessarily and, secondly, the annoying hiss that always accompanies escaping air is avoided. In broadcasting stations where quiet is essential, the avoidance of unnecessary noise is of course important.

The temperature of the used air is roughly 150°C and this is a temperature at which the air can often be put in to good use, e.g. for heating the building or, if the valve is employed in a generator for capacitive heating¹¹⁾, for pre-heating materials such as plywood etc.

Owing to the presence of the cooling fins, the weight of the TBL 12/100 is 30 kg net, which, although more than that of the TBW 12/100, is not too much for one person to handle.

Life

Theoretically, a valve of the type TBL or TBW 12/100 would reach the end of its life from one of the following causes:

- 1) Evaporation of the cathode, more especially of the thorium in the filament.
- 2) Alloying of the zirconium with the core material of the grid.
- 3) Saturation of the getters by liberated gases.
- 4) Deformation of components as a result of recrystallization, sagging of materials carrying relatively heavy mechanical loads, and so on.

Let us now examine these points one by one.

- 1) With cathodes of pure tungsten operating at a temperature of 2600°K , evaporation of the tungsten limits the life

to some thousands of hours. Thoriated tungsten cathodes work at 2000°K , and calculations show that it takes about 100 000 hours for all the thorium to evaporate from the filament.

¹¹⁾ See, for example, Philips tech. Rev. 11, 232-240, 1950.

- 2) Practical results have shown that after 10 000 working hours there is no sign of alloying of the zirconium on the grid with the core material, at any rate as long as the valve is operated within the limits specified. Tests, however, have shown that overloading can be very damaging in this respect, so that care must be taken not to exceed the maximum ratings.
- 3) The getters (Zr) are present in such ample quantities that there is no likelihood of their becoming saturated unless the valve develops a leak.
- 4) It has already been pointed out that in order to minimize the risk of physical distortion, all mechanically loaded components of the valve are made of metals which are not easily deformed, i.e. tungsten and molybdenum. Welded joints, which in the case of these metals are usually very brittle, have been avoided in the design wherever possible; for example all the joints in the grid are effected by riveting or by what is known as the "cut-squash" method.

Ultimately practical tests will have to prove whether the good results to be anticipated from these considerations will be achieved. So far, the results have been very satisfactory, both in broadcasting transmitters and generators, and mention may be made of the Mühlacker broadcast transmitter near Stuttgart (*fig. 8*), where a number of TBW 12/100 valves have been in use under full load since April 1950. The life so far is 12 000 hours and there have been no breakdowns. *Fig. 9* further shows a TBL 12/100 which has been working at the Westerglen (Scottish) transmitter of the B.B.C. since 1st November 1950, also without interruption.

Economy through higher efficiency

The high output efficiency of the TBL and TBW 12/100 — practically 80% as compared with 65 to 75% in older types — shows a saving of roughly 15 kW per valve on full load, a similar amount being saved by reason of the lower heater power required. On 5000 working hours per year, therefore, the amount of power saved is not less than 150 000 kWh per valve.

At the Westerglen station (50 kW) the introduction of the TBL 12/100 has increased the overall efficiency of the transmitter from 50% to 56%.

Summary. The paper describes two models of a new transmitting triode, the TBW 12/100 and TBL 12/100, the former with water cooling, the latter with forced air cooling. The filament is of thoriated tungsten, which has much better emission

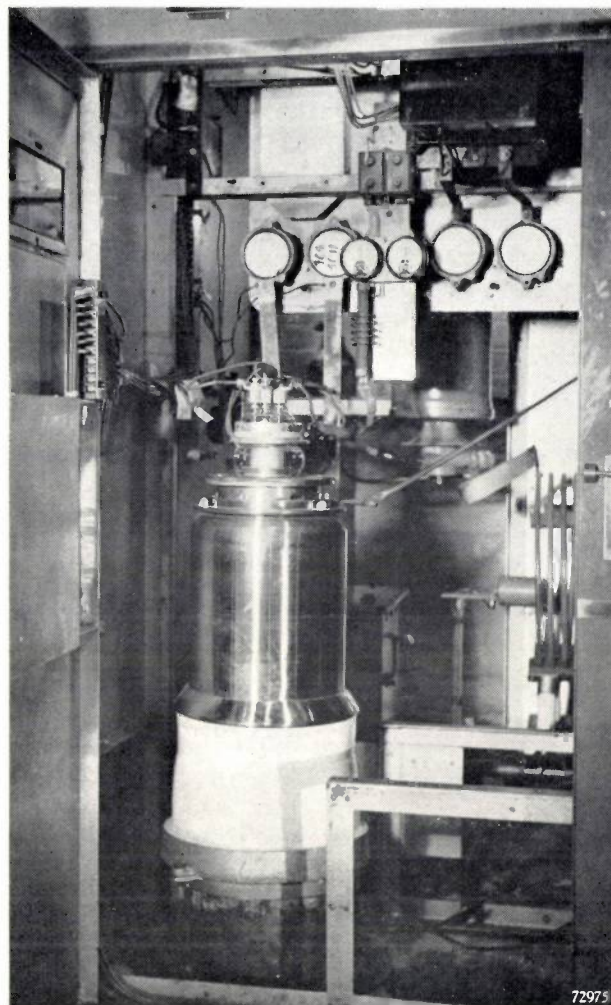


Fig. 9. A TBL 12/100 in the 50 kW Westerglen (Scottish) transmitter of the B.B.C.

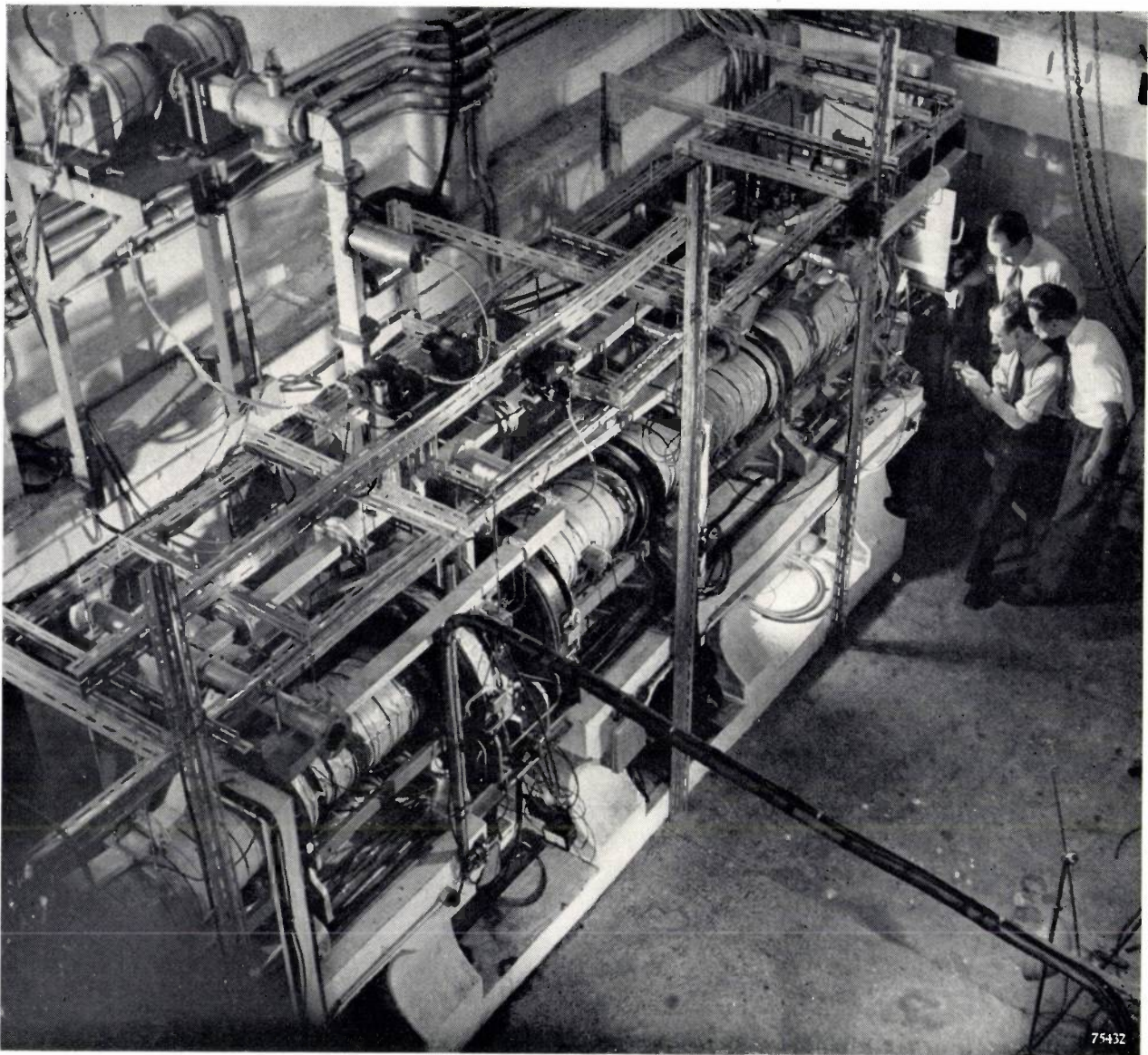
properties than pure tungsten, but could not be employed in high-power valves until more efficient getters had been developed. The new valves contain three getters, all of specially prepared zirconium and of which one functions slowly but has a high gettering capacity, one works rapidly, in the form of a layer of zirconium on the grid, and one is specifically for absorbing hydrogen.

The advantages as against comparable valves with pure tungsten filament are: no flashover (Rocky Point effect), thanks to the lower anode voltage required; lower filament consumption (3.5 kW); higher efficiency (a good 80% at frequencies up to 30 Mc/s, approx 60% at 68 Mc/s); longer life, easier replacement because of the lighter weight and smaller dimensions. With 80% output efficiency, an economy of 150 000 kWh per valve and per year of 5000 working hours can be achieved.

At frequencies up to 30 Mc/s the output power is 105 kW (telegraphy) or 65 kW (telephony). For water-cooling, the "Grip-o-matic" water jacket has been introduced, ensuring very high turbulence of the water and a water consumption of less than 1 litre/min per kW to be dissipated.

The air-cooled type is equipped with a very large number of cooling fins ("Multifin" system) of which each section is cooled by separate jets of air. With an air consumption of 1 m³/min per kW, 50 to 60 W per cm² of anode area can be dissipated. The used air (temperature 150 °C) may be utilised for heating or drying purposes.

15 MILLION ELECTRONVOLT LINEAR ELECTRON ACCELERATOR DURING CONSTRUCTION



Photograph Walter Nürnberg

One of the forthcoming issues of this Review will be largely devoted to a description of this electron accelerator, designed and built by a team of scientists at the Mullard Research Laboratories, Salfords, England, under the direction of C. F. Bareford and M. G. Kelliher. The installation was ordered by the Atomic Energy Research Establishment at Harwell and was handed over to them during August 1952, after several months of testing at Salfords.

The photograph pictures the accelerator, whose length is more than 6 metres, during trial assembly in an underground shelter built especially for this purpose at M.R.L. The 10 cm waves which impart their energy to the particles injected into the accelerator are produced by a magnetron which is just visible in the upper left hand corner.

THE NOISE OF ELECTRONIC VALVES AT VERY HIGH FREQUENCIES

II. THE TRIODE

by G. DIEMER and K. S. KNOL.

621.396.822:621.396.645.029.6

The noise of a triode at very high frequencies is determined (as is the noise of a diode) by the form of the current pulses produced in the external circuit by the individual electrons passing through the valve. If the triode is to function satisfactorily as part of an amplifier, the noise produced by the valve in the output circuit must be minimized. This can be achieved satisfactorily by a judicious choice of the elements in the external circuit.

Noise of a triode

Noise of a triode is caused, as in the case of a diode, by fluctuations in the state of motion of the electrons as they pass through the valve. As explained in a previous article¹⁾, the noise phenomenon can be investigated by resolving the convection noise current in the valve into its components at various frequencies (Fourier analysis), and investigating the relationship existing between these convection current components and the noise current components in the external circuit of the valve. The convection current is defined as the net amount of charge passing a given cross-section of the valve per second. At frequencies of the order of the reciprocal of the transit time of the electrons, the convection current is related only indirectly to the current in the external circuit, since the latter is actually an induced current (the electrons induce charges on the electrodes on their way through the valve). The current in the external circuit, or at least its Fourier components, can, however, be expressed in terms of the Fourier components of the convection current which starts at the cathode, by means of the transit time function Ψ . This function depends on the transit time τ of the electrons between the planes of two electrodes in the valve, and, of course, on the frequency f .

Similarly, if a signal voltage is present in the input circuit of the valve, it is possible to express the signal voltage in the output circuit in terms of the convection signal current produced by the signal, by means of a transit time function. This convection current passes, however, in an electric field which has a certain phase relation with the externally applied signal. There is obviously no such relationship for noise currents. As a consequence, the

transit time function for the signal currents differs from that for the noise currents.

In a triode two "interelectrode spaces" may be distinguished, namely the cathode-to-grid space and the grid-to-anode space (cg-space and ga-space respectively). In order to investigate the noise and signal currents, the convection currents in both spaces must be considered in turn. It will be seen that a relationship exists between the two currents, a relationship depending on the nature of the circuit in which the triode is employed.

Triode circuits

Three different types of circuit are possible for a triode, in which each electrode in turn is the common electrode of the input and output circuits. These three circuits are known as: grounded-cathode, grounded-grid and grounded-anode circuits. They are shown diagrammatically in *fig. 1*; 1 and 2 indicate input and output circuits respectively, which for the moment will be considered to be short-circuited for high frequencies, so that H.F. currents may flow in these leads but no H.F. voltages occur between the planes of the electrodes.

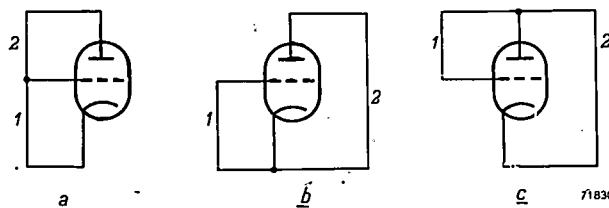


Fig. 1. Schematic representation of the three ways of connecting a triode: (a) grounded-grid circuit; (b) grounded-cathode circuit; (c) grounded-anode circuit; 1 is the input circuit, 2 the output circuit, both considered to be short-circuited for H.F. currents.

For low and medium frequencies, the grounded-cathode circuit is usually employed, whilst the grounded-grid circuit is preferred in amplifiers for very high frequencies. In the latter circuit the grid

¹⁾ G. Diemer and K. S. Knol, The noise of electronic valves at very high frequencies. I. The diode, Philips tech. Rev. 14, 153-164, 1952 (No. 6) This article will be referred to as I in the following.

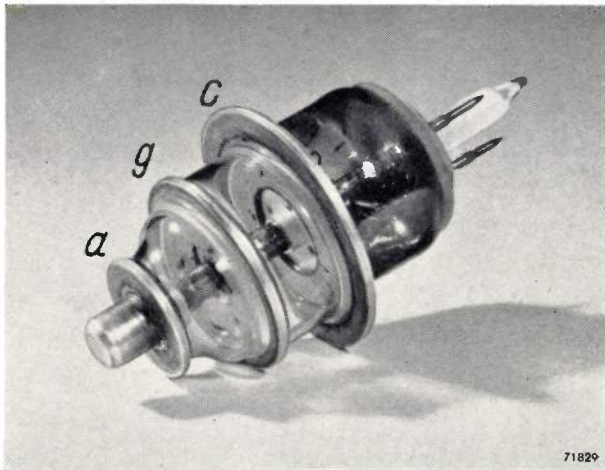


Fig. 2. Experimental disc-seal triode: *c* is the cathode disc, *g* the grid disc and *a* the anode disc (fitted with a cap).

acts (when it is earthed for high frequencies) as a natural screen between the input electrode (cathode) and the output electrode (anode). Moreover, in grounded-grid circuits, if the grid is mounted on a metal disc which extends outside the glass envelope ("disc-seal" triode, *fig. 2*), the H.F. input currents and the H.F. output currents are not able to influence each other, since these currents flow on opposite sides of the grid disc as a consequence of the skin effect. In this way the electric and magnetic couplings, which could cause feed-back, is made as small as possible. The grounded-anode circuit is seldom employed, and in the following it will be left out of consideration.

Current pulses in a triode

The potential distribution between the cathode and grid planes in a triode differ but little from that in a diode if the grid and anode voltages are not extremely high. There is a potential minimum located somewhere between cathode and grid. For the sake of simplicity, this potential minimum will be considered to be immediately adjacent to the cathode (as, in fact, it frequently is). An electron passing from cathode to grid (transit time τ_{cg}) produces, precisely as in the case of the diode, an almost parabolic current pulse in the external circuit, in this case the lead between cathode and grid; this is the input circuit in both grounded-grid and grounded-cathode circuits. This current pulse is shown graphically as curve 1 in *fig. 3a*. After passing the grid, the electron produces, during its passage from grid to anode (transit time τ_{ga}), a current pulse as shown by curve 2 in *fig. 3a* in the lead between the two electrodes, i.e., in the case under consideration, the output circuit. In the grounded-grid circuit the latter pulse flows only in this out put lead.

In the grounded-cathode circuit, however, the second pulse flows through the input and output leads in series, the current direction in the input circuit being opposed to that of the first pulse in this lead. Thus, whilst *fig. 3a* gives the current pulses in input and output circuits for grounded-grid circuits, the total current pulses in both circuits of the grounded-cathode circuit is as shown in *fig. 3b*.

The noise of a triode will now be derived from these current pulses.

The current pulses in the input lead are responsible for the noise in this lead, the so-called input noise. Similarly, the current pulses in the output circuit are responsible for the output noise. As will be shown later, it is not sufficient to study only the output noise; if the input circuit contains finite impedances, the input noise is also of importance in defining the properties of the valve and circuit. It will therefore be useful to discuss both input and output noise.

The two current pulse curves 1 and 2 in *fig. 3* obviously have an area equal to the electronic charge. If a Fourier analysis is made of the current pulses in the input lead of a grounded-cathode circuit, it is found that at low frequencies the two pulses practically neutralise each other, in other words, the amplitudes of the input components at low frequencies are practically zero. This is not the case for grounded-grid circuits; at low frequencies, however, the pulses may be considered to be infinitely small, giving for these frequencies a spectrum with

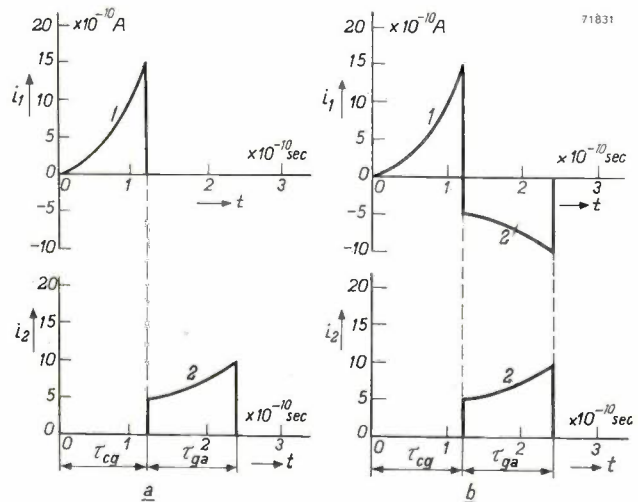


Fig. 3. (a) An electron moving from cathode to anode in a triode produces current pulses in the input circuit (i_1) and the output circuit (i_2). With the simplifying assumption that the potential distribution between cathode and grid is such that Child's law applies (see I), these pulses, in the grounded-grid circuit, have the shape 1 (in the input circuit) and 2 (in the output circuit). (b) In the grounded-cathode circuit, in addition to the pulse 1, a pulse 2 flows through the input circuit (but in the opposite direction). Only pulse 2 flows in the output circuit.

a finite amplitude, independent of frequency (see I). This is true also for the output currents in both circuits. In other words, at low frequencies the output noise in grounded-cathode circuits, and both input and output noise in grounded-grid circuits, are directly comparable with the normal shot effect in diodes. The input noise in the grounded-cathode circuit is practically zero.

We proceed to investigate noise at frequencies of the order of the reciprocal transit time (or less). If the input and output circuits are still considered as being short-circuited, there exist certain phase relationships between the noise currents in the input and output leads. First, the currents in the input lead will be investigated by means of a vector

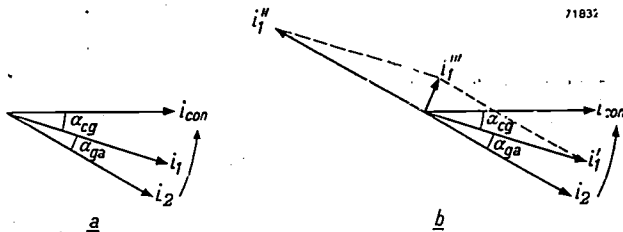


Fig. 4. Vector diagram for the noise currents in a triode. The vectors of the current components at a given frequency are shown, the curved arrow points in the direction of rotation, the angle between two vectors is the phase difference of these current components. i_{con} represents the convection noise current of the electrons which start their trajectory through the valve at the cathode; in drawing the amplitude of i_{con} , account has been taken of the noise suppression due to the space charge; i_1 is the noise current in the input lead of the triode, i_2 that in the output lead. The angle α_{cg} between i_1 and i_{con} is proportional to the transit time τ_{cg} in the cathode-to-grid space, the angle α_{ga} between i_2 and i_1 is proportional to τ_{ga} . (a) applies to grounded-grid circuits; (b) to grounded-cathode circuits. In the latter circuit the current in the input circuit consists of two components i_1' and i_1'' corresponding to the pulses 1 and 2 in fig. 3b; i_1' lags i_{con} by the phase angle α_{cg} , i_1'' is 180° behind i_1' (the current in the output circuit) and has the same amplitude as i_1' . The total noise current in the input circuit is i_1''' . For low frequencies α_{ga} is almost zero and i_1''' is then also zero. Only at high frequencies has i_1''' a finite value, albeit small, and then leads practically 90° with respect to i_1' .

diagram (fig. 4). The arrows in this diagram represent the components of the various noise current components at a given frequency f .

In grounded-grid circuits the phase of the noise current i_1 in the input lead lags by a certain angle, the transit time angle α_{cg} (related to τ_{cg} and the frequency by a transit time function), behind the convection current i_{con} which starts at the cathode (the potential minimum). The noise current i_2 in the output lead lags behind i_1 by a transit time angle α_{ga} . The resultant noise current in the input lead is in this case given directly by i_1 , the magnitude of which is almost independent of frequency for medium frequencies.

The case of the grounded-cathode circuit is different. The vector diagram corresponding to this

circuit is given in fig. 4b. Firstly, a noise current i_1' flows in the input lead lagging in phase by an angle α_{cg} behind i_{con} , and, secondly, a noise current i_1'' which is in anti-phase with the noise current i_2 in the output lead. The latter lags by a further angle α_{ga} behind i_1' . The resultant noise current in the input lead is thus i_1''' . A comparison of figures 4a and 4b clearly shows that in grounded-cathode circuits the resultant noise current in the input lead is much smaller than in grounded-grid circuits. At low frequencies the input current in grounded-cathode circuits is, as previously shown, almost zero. (This is because α_{ga} is practically zero at low frequencies.) It is true that at high frequencies a finite input noise current occurs, but this is still quite small. It is then seen that even at high frequencies the pulses 1 and 2 of fig. 3b cancel one another to some extent.

The contribution of the shot effect to the noise current at high frequencies in the input lead of a triode is called the induced grid noise (or the transit time noise) of the triode²⁾. It has been shown that this noise is much smaller for grounded-cathode circuits than for grounded-grid circuits. These induced grid currents have no influence on the output noise currents, so long as the input is short-circuited for H.F. currents. The influence at finite values of the input admittance is dealt with below. It will be seen that this influence is not always harmful.

If an H.F. signal voltage V_g is applied to the grid of a triode, a further vector diagram may be constructed for the Fourier components of the signal currents corresponding to the frequency of the voltage applied, now however with transit time angles: $\alpha'_{cg} \neq \alpha_{cg}$ and $\alpha'_{ga} \neq \alpha_{ga}$, since the transit time functions are different (fig. 5). As can be seen, the component I_1'''' of the input current which is in phase with the input signal is much smaller in the case of grounded-cathode circuits than in the case of grounded-grid circuits. This means that the electronic input damping, at least for medium frequencies, is smaller in the former case. At frequencies of the order of the reciprocal transit time, the damping in the case of grounded-cathode circuits increases. This is known as transit time damping.

This transit time damping may be regarded as an fictitious noise source producing the transit time noise (i.e. the high-frequency noise in the input circuit). In this case the equivalent temperature of the "transit time resistance" appears to be about

²⁾ This is an extension of the similarly named concept introduced by C. J. Bakker: Philips tech. Rev. 6, 129-138, 1941.

1.5 times the cathode temperature. It is, however, unsound to introduce the concept of such a transit time resistance, since the phase relations between the noise currents in the input and output circuits are then neglected. Such a correlation between

Looked at in this way, it would appear that the noise in the output circuit is determined by the α - β -electrons. From the above, however, it will be clear that this noise is caused almost wholly by the much larger number of electrons which always reach the anode (called β -electrons in Part I). If only these β -electrons are considered together with their influence on the potential minimum, the output noise can be computed without the mutual conductance being taken into consideration. The output noise, therefore, depends only in a very indirect way on the mutual conductance, and on the equivalent noise resistance. It is therefore apparent that, if the noise is defined by means of such a resistance, the relationship between the various causes of the noise is very easily overlooked.

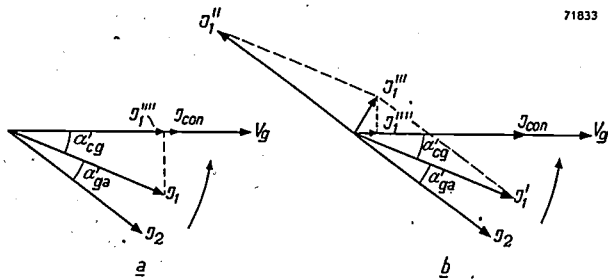


Fig. 5. Vector diagrams for the signal currents in a triode. The symbols have the same meanings as in fig. 4 (to show that signal currents are now referred to, the small letters are replaced by capitals and the transit time angles are dashed). The convection current which starts at the cathode, i_{con} , is in phase with the signal voltage V_g . (a) Grounded-grid circuit. In the latter circuit the component I_1'''' , which is in phase with the signal voltage (and thus with I_{con}) is very small: there is only slight electronic input damping.

The triode in amplifier circuits; noise factor

input and output currents would apply fully only if all the electrons had the same transit time. Due to the irregularities in the cathode surface, the velocity distribution of the electrons, and the inhomogeneity in the potential distribution introduced by the grid wires, there appears, however, a spread in the transit time. The correlation is then no longer complete, but does not wholly disappear³⁾.

Attention will now be paid to the influence that the electronic noise of a triode has on the useful gain that can be given by the valve; when used in an amplifier stage, the valve circuits thus not being short-circuited. First, the concept of useful gain will be considered. A measure of the relative influence of the noise is given by the signal-to-noise ratio, by which is meant the ratio of the signal power to the noise power, measured in the direction from the input to the output through a given "cross-section" of the amplifier (e.g. the output terminals).

For the same reason, it is undesirable to make use at high frequencies of the concept of equivalent noise resistance, i.e. a resistance that is imagined to exist in the input lead of a hypothetical noise-free valve, and that is just high enough to produce at room temperature the same noise currents in the output lead as are in fact produced by the valve itself. It can be shown that this fictitious noise resistance is closely related to the mutual conductance of the valve.

The possibility of obtaining effective gain of a given small signal is entirely dependent on the signal-to-noise ratio. If the signal is so weak that the signal-to-noise ratio at the output terminals is smaller than unity, the output signal is "drowned" by the noise and becomes very difficult or impossible to separate from the noise. In practice, the tolerable signal-to-noise ratio at the output side of the amplifier depends on the nature of the application. For television receivers, for example, a ratio of the order of 100 is desirable, whilst for radar apparatus a signal-to-noise ratio of just over 1 is considered satisfactory. In special amplifiers, as used, for example, for measuring noise from inter-stellar space at decimetric waves, one must be satisfied with signal-to-noise ratios of the order of 1%. It has been found, however, that by special measuring procedures, variations in the average value of the "signal" to be determined (which in this case has itself a noise character) of approximately 0.1% of the amplifier noise can still be detected.

The mutual conductance is determined by those electrons that leave the cathode with forward velocities falling within a narrow velocity range of such values that their behaviour (passing the potential minimum or not, and thus contributing to the anode current or not) can be entirely determined by the variations in depth of the potential minimum, such as may be caused by grid voltage variations. In Part I these electrons were termed α - β electrons (they are also responsible for the static internal resistance of a diode).

At very high frequencies, the signal-to-noise ratio at the output of the amplifier is affected by the

³⁾ A. van der Ziel and A. Versnel, Philips Res. Rep. 3, 255-270, 1948; W. Kleen, Frequenz 3, 209-216, 1949; A. van der Ziel, Can J. Technology 29, 540-553, 1951.

circuit elements external to the valve⁴). This is not the case at low frequencies. The cause can be found in the different behaviour of the signal currents and noise currents on their way through the valve, as previously mentioned. If there is an admittance (such as is caused by the presence of feedback) in the circuit between input and output terminals, any noise voltage existing at the input terminals can cause a corresponding noise voltage at the output terminals via two different paths, i.e. through the valve itself and by-passing the valve. A signal in the input similarly makes two contributions to the signal in the output circuit. But, at high frequencies, the signal currents and noise currents are, on their way through the valve, affected in different ways (different transit time functions). Thus, variation of the feedback admittance will change the signal-to-noise ratio at the output terminals.

This point will be reverted to presently, but it might be remarked here that there is a possibility of minimising the noise in the output lead, and thus making the signal-to-noise ratio at the output terminals very large indeed. This may be achieved, for example, by so choosing the feedback admittance that advantage may be taken of the phase relations of the various noise currents. This possibility (which is almost always denied by current opinion) exists for both grounded-grid and grounded-cathode circuits.

The influence of the feedback admittance on the noise in the output circuit due to the phase relations existing between the various noise currents, emphasizes once again the undesirability of describing the H.F. output noise by the concepts of equivalent noise resistance and noise temperature.

The signal-to-noise ratio at the output terminals of the amplifier depends on the input signal. The latter may be varied over a wide range, and thus also the signal-to-noise ratio, without necessarily any change in the noise. In order to determine quantitatively the sensitivity of an amplifier (defined in this case as the weakest signal that may still be "usefully" amplified) the signal-to-noise ratio at a standard level of the input signal power should be known. The signal power which is equal in strength to the maximum noise power that can be produced by an arbitrary resistance in thermal equilibrium with the surroundings (i.e. $kT\Delta f$, see equation (2) in I) may, for example, serve this purpose; such a resistance is always present in the input circuit. The signal-to-noise ratio corresponding to this standard input signal is known as the noise

factor F ⁵). The noise factor is frequently expressed in decibels, and is then known as the noise figure N .

In the literature definitions of F are often seen which, at first sight, seem to differ one from the other. In order to conform as closely as possible with modern views, the above definition will be generalised (without, however, becoming at variance with it), as follows: F is the quotient of the signal-to-noise ratio at the terminals of the input resistance at room temperature — the latter being considered to produce pure thermal noise — and the signal-to-noise ratio at the output terminals of the amplifier. The previous definition is a special case of this definition: i.e. the standard signal was chosen so that the signal-to-noise ratio at the terminals of the input resistance was equal to unity.

Once the noise factor of an amplifier, defined in this way, is known, it is possible to compute directly, without further reference, how small the input signal to be amplified may be; if the signal-to-noise ratio is not to drop below the limit specified for the case under consideration. At not too high frequencies, the noise figure of a good radio or television amplifier must not exceed a few decibels.

As long as the amplifier amplifies linearly, F is independent of the amplitude of the signal. In the determination of F , the signal is frequently increased to such an extent that the signal-to-noise ratio at the output of the amplifier is equal to unity. In this case F is equal to the ratio of the noise power fed to the input to the noise power of the (thermal noise producing) input resistance. If all the noise is imagined to be produced in an input resistance which is supposed to have a given, fictitious temperature, this temperature must be equal to F times the room temperature. For an ideal amplifier, F should be equal to unity.

At high frequencies great difficulties are encountered in the measurement of F . For example, even in a single stage amplifier, the noise produced by the measuring instrument used to determine the signal-to-noise ratio at the output terminals influences the determination of F . Again, the measured thermal noise of a resistance depends on the bandwidth of the amplifier. Further, in the case of super-heterodyne amplifiers (in agreement with the definition) when considering the bandwidth, it is necessary to take into account also the image frequencies in the frequency characteristic as passed by the input of the amplifier at the input side. If this is done, the noise factor is then roughly twice as great as when the image frequencies are left out of consideration (which is in better agreement with the sensitivity found in practice).

It can easily be shown, if the signal source is not matched to the input of the amplifier (that is, if the internal resistance of the signal source differs from the input resistance of the amplifier), and the power supplied by the signal source is thus not at its maximum, that the correct value of F can be calculated by using in the definition that value of the input signal-to-noise ratio which would apply if the matching were correct.

⁴) K. S. Knol and A. Versnel, *Physica* 15, 462-464, 1949.

⁵) H. F. Friis, *Proc. Inst. Radio Engrs.* 32, 418-423 and 729, 1944.

The noise level reached in the output circuit of an amplifier when finite admittances are present in the valve circuit will now be investigated, and at the same time the possibility, already mentioned, of achieving a low noise factor will be considered in greater detail. Only a few examples will be dealt with.

Amplifier with grounded-grid circuit, without feedback admittance

Fig. 6 shows an amplifier stage with a grounded-grid circuit. The valve admittances are separated into electronic admittances (with suffix *e*, indicated by broken lines) and non-electronic admittances, due, for example, to resistance losses in the leads. For the present, non-electronic feedback admittance will be assumed to be absent. First the dependence of the output signal on the input matching will be investigated. The signal source (aerial) is inductively coupled (ratio of transformation *t*) to the input circuit, and supplies to it an electromotive force *tV*. The input admittance due to the coupling with the signal source is equal to g/t^2 , in which *g* is the conductance of the source; this is in series with the impedance Y_1 . The admittance between grid and anode is denoted by Y_2 . It can be shown that the available signal power at the output increases with *t* (in spite of the fact that the input voltage

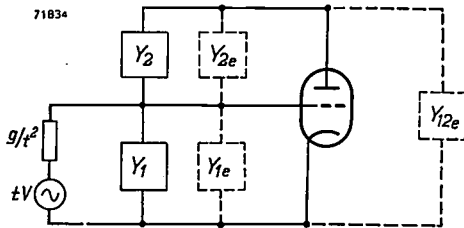


Fig. 6. Basic circuit of an amplifier stage with grounded-grid circuit without feedback of a non-electronic nature. The various admittances are split into electronic (drawn in broken lines and with index *e*) and non-electronic admittances; Y_1 and Y_{1e} are the input admittances, Y_2 and Y_{2e} output admittances, Y_{12e} electronic feedback admittance.

then decreases), since in grounded-grid circuits the output admittance depends largely on t^6). The output noise current may be studied as follows, by means of a vector diagram (fig. 7). Two types of convection currents originate from the cathode (potential minimum), i.e. the primary noise currents, described by $2eI_a \Delta f \Gamma^2$ (see I, formula (6) where Γ is the noise-suppression factor of the space charge) and secondary noise currents, resulting from induced fluctuations ΔV_g in the grid voltage.

These latter are caused by the noise currents flowing in the external input admittance: $\Delta V_g = -i_1/(Y_1 + g/t^2)$. The total convection noise current originating at the cathode is represented by the vector i_{con} . The

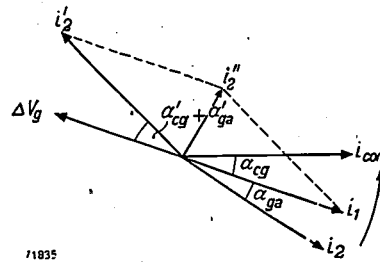


Fig. 7. Vector diagram of the noise currents in the output circuit of a triode in a grounded-grid circuit. The fluctuations of the grid voltage ΔV_g , which are in anti-phase with the input noise current i_1 , produce a secondary output noise current i_2' , which at not too high frequencies can partially compensate the primary output noise current i_2 (resultant current i_2''). i_1 is the input noise current, i_{con} the convection current which starts at the cathode.

noise current i_1 in the external input circuit lags behind i_{con} by the transit time angle α_{cg} , and ΔV_g is in its turn in anti-phase with i_1 if Y_1 is taken to be real and g/t^2 is small (i.e. for large values of *t*; in this case the most favourable noise suppression is obtained). The output noise current is made up of i_2 (lagging behind i_1 by the angle α_{ga}) and the secondary noise current i_2' (caused by ΔV_g), which lags behind ΔV_g by the angle $\alpha_{cg}' + \alpha_{ga}'$. Herein α_{cg}' and α_{ga}' are "signal" transit angles, since i_2' is caused by ΔV_g , so that it is affected in the same way as a signal current. At low frequencies, i.e. at medium large transit time angles, the phase difference between i_2 and i_2' can be made practically equal to 180° (α_{cg}' and α_{ga}' then do not differ greatly from α_{cg} and α_{ga} respectively), so that the resultant output noise current $i_2'' = i_2 + i_2'$ can become almost zero at high ratios of transformation ⁷⁾. The noise factor is then at its minimum value. This is illustrated in fig. 8. The measurement of *F* as a function of the ratio of transformation *t* is not carried out in practice by placing a transformer between the aerial and the input circuit, but by shunting a resistor R_a equivalent to the reflected aerial resistance, i.e. t^2/g , directly across the input circuit. In fig. 8, *F* is plotted against the value of this resistor for the case of an experimental short-wave triode. The measurements were carried out at a wavelength of 7 m. The very low minimum value of *F* is reached when the reflected aerial resistance

⁶⁾ A. van der Ziel, Philips Res. Rep. 1, 381-399, 1946.

⁷⁾ A. van der Ziel and A. Versnel, Philips Res. Rep. 3, 255-270, 1948; A van der Ziel and K. S. Knol, Philips Res. Rep. 4, 168-179, 1949.

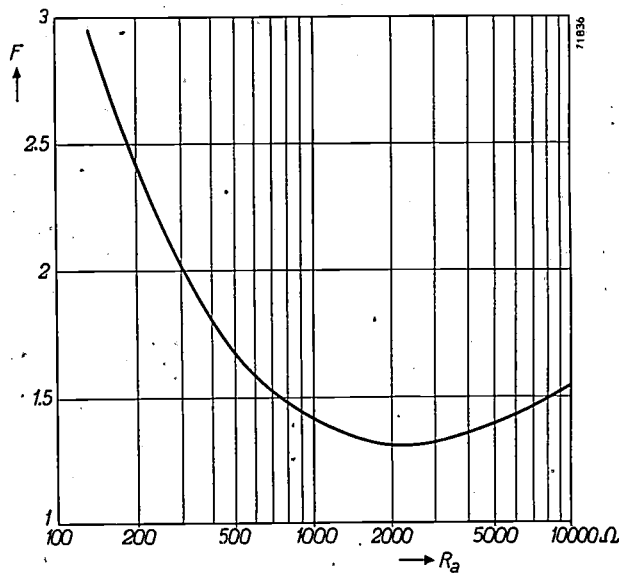


Fig. 8. Noise factor F of an experimental triode in grounded-grid circuit as a function of the reflected aerial resistance R_a , measured at a wavelength of 7 m. F is at its minimum when R_a is about 2000 Ω . With an aerial resistance of 70 Ω this corresponds to a ratio of transformation of 5.4.

is about 2000 Ω . At an aerial resistance of 70 Ω this corresponds to a ratio of transformation of $t = 5.4$.

At centimetric wavelengths a number of effects appear which cause F to increase. These effects are as follows:

- 1) The total-emission noise (i.e. the noise originating in the input circuit due to the electrons which recoil before the potential minimum, the so-called α -electrons, see I) commences to make an important contribution to the induced input noise.
- 2) The total-emission damping, i.e. the damping caused by the α -electrons, reduces the gain.
- 3) The phase relations between the various noise currents become such that hardly any compensation exists (very large transit angles).
- 4) The clearance between grid and cathode in valves used for centimetric waves is frequently so small that the potential minimum lies close to the grid, so that the controlling action of the grid wires decreases. The spread in transit times becomes much larger, since the distances between the grid wires are no longer small in comparison with the clearance between grid and cathode.
- 5) The feedback between the output and input circuits increases, in general, strongly⁸⁾.
- 6) The valve conductances increase, due to increases in skin-effect losses and dielectric losses.

⁸⁾ S. D. Robertson, Bell Syst. Techn. J. 28, 647-655, 1949; G. Diemer, Philips Res. Rep. 5, 423-434, 1950.

This results in a reduction of the gain⁹⁾.

By using a special technique in the manufacture of triodes, it is possible to achieve very small inter-electrode clearances and high current densities. The effects mentioned above are then largely overcome. In this way noise figures in the centimetric wave range can be obtained which are as low as those for normal broadcast valves for metric waves. Philips triode EC 80, of fairly normal construction and pin leads¹⁰⁾, has a noise figure of 7 dB at $\lambda = 2$ m. With an experimental disc-seal triode such as that shown in fig. 2, it is possible to reach at $\lambda = 10$ cm a noise figure of about 10 dB with a power gain of about 10 dB¹¹⁾. At $\lambda = 75$ cm, it was shown that the noise figure in a grounded-grid circuit is larger than that in a grounded-cathode circuit¹²⁾. This result indicates that for extremely low noise figures the grounded-cathode circuit is to be preferred, provided the gain can remain at a satisfactorily high value and there is no risk of the amplifier becoming unstable.

These very low noise figures can be reached by taking advantage of the correlations that exist between the various noise currents. The influence of a feedback admittance in an amplifier with grounded-cathode circuit will now be discussed, these correlations being taken into account.

Amplifier with grounded-cathode circuit and non-electronic feedback admittance

The output of an amplifier with a grounded-cathode circuit is assumed to be short-circuited for H.F. currents (fig. 9). By means of a vector diagram once

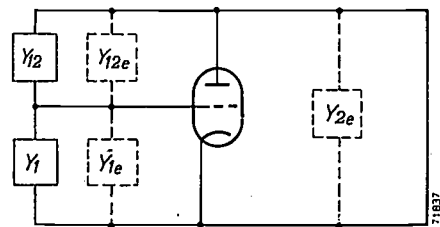


Fig. 9. Basic circuit of an amplifier stage with grounded-cathode circuit and finite feedback admittance, with short-circuited output. The symbols have the same meanings as in fig. 6.

⁹⁾ K. Rodenhuis, Philips Res. Rep. 5, 46-77, 1950; G. Diemer and K. Rodenhuis, Philips Res. Rep. 7, 36-44, 1952 (No. 1).

¹⁰⁾ A description of this valve may be found in: K. Rodenhuis Two triodes for reception of decimetric waves, Philips tech. Rev. 11, 79-89, 1949.

¹¹⁾ In a provisional report (Philips Res. Rep. 5, 153-154, 1950) 7 dB was given as the lowest value of N . These measurements proved later to contain some inaccuracies.

¹²⁾ These measurements were carried out by Dr. Stumpers of this laboratory.

again, the behaviour of the noise currents (fig. 10) will be studied. Consider a small convection current i_{con} originating at the cathode; this current represents the resultant of all the convection currents contributing to the noise. The current primarily

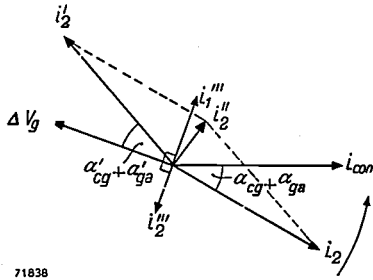


Fig. 10. Vector diagram of the noise currents for a triode in a grounded-cathode circuit. i_{con} is the noise current which starts at the cathode. This results in a primary noise current i_2 in the output lead. The resultant noise current i_1''' in the input lead produces grid voltage fluctuations ΔV_g , which, in turn, cause noise currents i_2' in the output circuit. The resultant output noise current is i_2'' . If a capacitive feedback admittance is present, this results in further noise currents i_2''' . By suitable choice of input and feedback admittances it is possible to make i_2'' and i_2''' largely compensate each other.

excited in the output circuit is i_2 ; because of the input admittance Y_1 , in which for convenience the reflected aerial resistance $R_a = t^2/g$ is included, there is a small resultant noise current i_1''' which leads i_2 in phase by approximately 90° (see fig. 4b). This is quite different from the case of the grounded-grid circuit, in which the resultant noise current in the input circuit leads i_2 by the small angle α_{ga} only (fig. 7). The noise current i_1''' produces a grid

voltage variation $\Delta V_g = -i_1'''/Y_1$; the phase of ΔV_g is controlled with respect to i_1''' by the choice of Y_1 . The most favourable noise suppression is obtained when Y_1 is purely capacitive (thus t must again be large). The phase difference between ΔV_g and i_1''' is then 90° , and the difference between ΔV_g and i_2 is almost 180° . This is an analogous situation to that reached with a grounded-grid circuit when a real value was chosen for Y_2 . The grid voltage variation produces a further current in the output circuit lagging almost as much in phase behind ΔV_g as i_2 lags behind i_{con} , i.e. the angles $(\alpha_{cg}' + \alpha_{ga}'')$ and $(\alpha_{cg} + \alpha_{ga})$ respectively. For medium transit angles, just as with a grounded-grid circuit, it is possible also with a grounded-cathode circuit, by suitable choice of Y_1 , to make the angle between i_2 and i_2' almost equal to 180° and to give the two currents almost the same amplitude, so that the resultant output noise current i_2'' is very small. With the compensation thus achieved the gain is also decreased (the matching of the input circuit to the aerial is less favourable since t had to be $\gg 1$), but since the noise decreases relatively more quickly, the noise factor decreases. This is illustrated in fig. 11a, which relates to an EF 50 pentode used as a triode at a wavelength of 17 m. The feedback admittance which is a consequence of the capacitance between grid and anode is neutralised by shunting this capacitance by an inductance. The feedback admittance is thus reduced almost to zero. Curve 1 shows the noise factor F as a function of

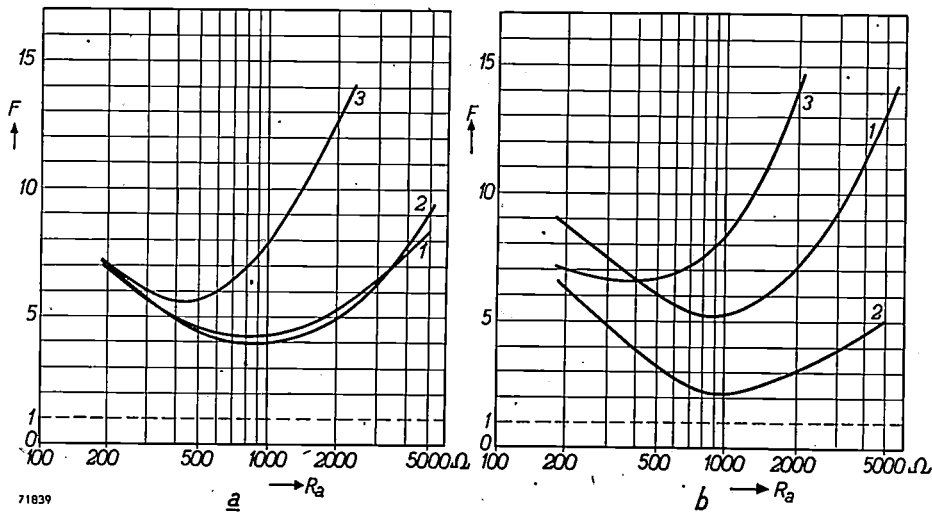


Fig. 11. Noise factor F measured for an EF 50 pentode connected as a triode in a grounded-cathode circuit at a wavelength of 7 m, as a function of the reflected aerial resistance R_a in the input circuit. (a) applies to the case where no non-electronic feedback admittance is present in the circuit, (b) for the case where this is present, i.e. when the capacitance between grid and anode is not neutralised. Curves 1: tuned input circuit, i.e. with real value of the input admittance Y_1 ; curves 2: capacitively detuned input circuit (Y_1 practically wholly capacitive); curves 3: inductively detuned input circuit (Y_1 practically wholly inductive). The noise factor always has a minimum at certain values of the reflected aerial resistance. The minimum is smallest for non-neutralised feedback and capacitive detuning (curve 2 of b).

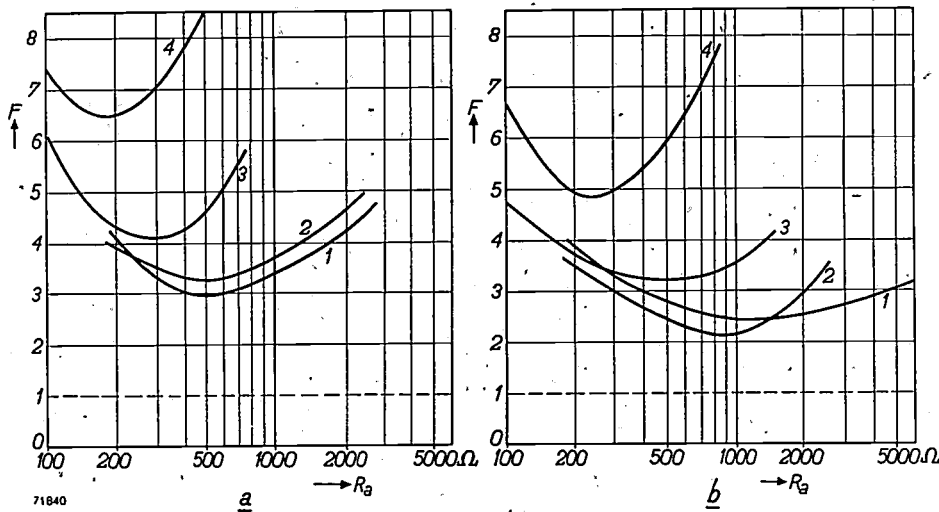


Fig. 12. Noise factor F , measured for an EF 42 pentode connected as a triode in grounded-cathode circuit at a wavelength of 7 m, as a function of the reflected aerial resistance R_a . (a) applies to a tuned input circuit (input impedance Y_1 real), (b) applies to an input circuit detuned by 5 pF (Y_1 practically wholly capacitive). Curves 1: no feedback admittance (neutralised); curves 2, 3 and 4: capacitive feedback with $C_t = 0.5, 1, \text{ and } 1.5$ pF respectively. The minimum noise factor is lowest with capacitive detuning and 0.5 pF feedback admittance (curve 2 of b).

the reflected aerial resistance R_a with tuned input circuit (Y_1 real). There is again, as with the grounded grid circuit, a minimum noise figure, in this case at $R_a \approx 1000 \Omega$. Curve 2 gives the variation of the noise as a function of R_a for a capacitively detuned input circuit (Y_1 capacitive apart from the small component i^2/g), and curve 3 for an inductively detuned input circuit (Y_1 practically wholly inductive). It is seen that, in agreement with the above, the smallest minimum noise factor is obtained with a capacitively detuned input circuit; inductive detuning increases the noise factor considerably (and also shifts the minimum).

The noise factor can be still further reduced, as will be finally shown, by making use of a feedback admittance Y_{12} between anode and grid. If this is made purely capacitive, ΔV_g produces a further noise current i_2''' in the output circuit, leading 90° with respect to ΔV_g ¹³). The sum of all the noise vectors in the output circuit is thus further reduced. An example of this is given in fig. 11b, which concerns the same valve as in fig. 11a. Now, however, neutralisation is omitted, so that a capacitive feedback is, indeed, present between anode and grid. The curves 1, 2 and 3 of fig. 11a now change to the corresponding curves 1, 2, and 3 in fig. 11b. It is clear that the presence of the feedback and the use of capacitive detuning does in fact markedly reduce the minimum noise factor compared with that without feedback.

According to fig. 10 it should therefore be possible, theoretically, to build a noise-free amplifier. A fuller examination shows, however, that with an optimum adjustment the signal gain drops below unity. Moreover, the correlation between the various noise currents is smaller in practice than it is in theory. It is clear, however, that a quite considerable improvement in noise factor is attainable.

¹³) See, for example, the references cited in notes ³) and ⁶).

In conclusion, fig. 12 gives measurements of the noise factor with an EF 42 pentode used as a triode at a wavelength of 7 m. In this case not only is the tuning of the input circuit varied, but also the feedback capacitance; this was effected by shunting various capacitors C_t across the neutralising inductance. At $C_t = 0$, there is no feedback admittance; values of C_t differing from zero produce a corresponding feedback capacitance. Fig. 12a applies to an amplifier with tuned input circuit (Y_1 real), and fig. 12b to a capacitively detuned input circuit (by 5 pF) (Y_1 practically wholly capacitive). The capacitive detuning of the input circuit clearly has a favourable influence on the noise factor. A small capacitive feedback admittance gives still further improvement; a large capacitive feedback has the opposite effect, however, in the same way as an inductive feedback admittance (not shown in the figure). From these figures it is seen that a suitable choice of feedback admittance is of great importance for rendering the noise factor as small as possible. This influence of the feedback admittance in grounded-cathode circuits is often denied in current opinions.

Summary. This article, which is a continuation of an article with the same title published recently, deals with the noise of a triode at very high frequencies. It is shown that in the input circuit of a triode, at frequencies of the order of the reciprocal of the transit time of the electrons, a new type of noise is produced, the induced input noise. This is caused by the fact that noise currents which are induced in the external circuit leads give rise to grid voltage fluctuations. The induced input noise may influence the output circuit noise if the valve circuits have finite impedances. A triode can be used in several types of circuit, e.g. in grounded-grid and in grounded-cathode circuits. In both types of circuit a certain amount of compensation may occur in the various noise currents at the valve input, since phase relations exist between the noise currents in both inter-electrode spaces. The behaviour of triodes as amplifiers in both circuits mentioned is discussed. The sensitivity of the amplifier depends on the noise in the output circuit, and is expressed by a specially defined noise factor.

The influence which a feedback admittance between input and output of the valve may exert is illustrated.

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

MEASURING METHODS FOR SOME PROPERTIES OF FERROXCUBE MATERIALS

by C. M. van der BURGT, M. GEVERS and H. P. J. WIJN. 621.318.134: 621.317.4

Ferroxcube materials differ from other soft ferromagnetic materials in that they are especially suitable for use at high frequencies. For this reason, in the measurement of their characteristics, methods are applied which for the greater part belong to those commonly used in radio- and radar-technique, and are thus rather unusual in the field of ferromagnetism.

Introduction

The technically most important properties of a soft ferromagnetic material are the permeability and the losses in an alternating field. These quantities, for Ferroxcube materials, have been amply dealt with in an earlier article in this Review¹). In this article, the discussion of the two quantities was divided into two sections, which dealt with the behaviour in weak and in strong fields respectively. The reason for this separation was principally that the magnetization process, for sufficiently weak fields, may be considered to be fairly closely reversible. In other words, the hysteresis phenomenon has in this case no noticeable influence and makes therefore only a negligibly small contribution to the losses. The behaviour in weak fields is thus easily understood, and the entire behaviour below the frequency range in which ferromagnetic resonance occurs, insofar as permeability and losses are concerned, is described by a complex material constant, the complex initial permeability

$$\mu_i = \mu_i' - j\mu_i'' \quad \dots \quad (1)$$

(in fact there are thus two real quantities μ_i' and μ_i'' requisite for a full description). As was amply described in I, the behaviour in strong fields is considerably more complicated, since the real part of the permeability and the losses at all fre-

quencies depend both on the frequency and on the field strength. In that case the properties mentioned must be derived from the magnetization curves (I, figs 7 and 8), the curves for the total losses (I, fig. 9) and the curves for the distortion.

The determination of the permeability and losses of metallic ferromagnetic materials is more complicated, as compared with Ferroxcube, by the fact that, as a consequence of their low resistivity, a high-frequency magnetic field has only a small depth of penetration (skin-effect): as a result, a certain effective permeability is measured, which is dependent on the shape and dimensions of the piece of metal. If it is desired to derive from this "effective" permeability the "true" permeability, which is a true material property independent of the shape, the skin-effect must be introduced quantitatively into the calculations, which naturally involves a certain amount of complication.

The occurrence of the skin effect in metals has the consequence that the losses are chiefly due to eddy currents. The completely different situation with regard to losses which is found when dealing with Ferroxcube has already been amply discussed in I, as also in a later article dealing with the applications of Ferroxcube materials²).

In the present article we will give a review of the methods available for the measurement of various

¹) J. J. Went and E. W. Gorter, Philips tech. Rev. 13, 181, 1952 (No. 7), referred to in the following as I. The system of rationalized Giorgi units will be used also in the present article. The permeability of the vacuum is then $\mu_0 = 4\pi \cdot 10^{-7}$ H/m.

²) W. Six, Philips tech. Rev. 13, 301, 1952 (No. 11), referred to in the following as II.

properties of Ferroxcube materials, such as those mentioned above, when subjected to a purely sinusoidal, alternating field. As in I, a distinction will be made between measurements in weak fields and measurements in strong fields. For both cases further consideration will be given to the methods of measuring the permeability and losses. Finally, we shall also discuss the measurement of the magnetization curve and the distortion, the importance of which was clearly shown in II.

In all these considerations we shall refrain from going into details of the measuring equipment, but limit ourselves to the examination of the relative merits of the different methods available. It will frequently be necessary to employ different methods for the determination of the same quantity, according to the frequency and the field strength at which the measurements are to be made, and the order of magnitude of the quantity to be measured in the given circumstances.

Recapitulation of certain concepts

Permeability and losses in weak fields

To obtain a better insight, we shall note here the most important definitions bearing on the concepts of permeability and losses. For a fuller discussion the reader is referred to I and II.

If Ferroxcube is employed as core material in an ideal toroid, and if a weak uniform magnetic field H is excited by means of a weak current, passing through the toroid and varying purely sinusoidally with time, then the flux density B is nearly enough a linear function of the field H . As a consequence of the losses occurring in the ferromagnetic material (in this case, to a close approximation, exclusively "residual losses"), a phase difference appears between B and H . The relation between B and H is given by $B = \mu_i H$, in which μ_i is the previously mentioned, complex, initial permeability (formula 1). The phase angle δ_F between B and H is now given by

$$\tan \delta_F = \frac{\mu_i''}{\mu_i'} \dots \dots \dots (2)$$

As was amply described in II, this method of description amounts to replacing the coil and core by a loss-free coil with inductance L_s and a resistance R_s connected in series. For an ideal toroid of length l (measured along the centre line), cross-section O and number of turns n , which closely surrounds a core of ferromagnetic material with (complex) initial permeability μ_i , the following equations apply:

$$R_s = \omega n^2 \frac{O}{l} \mu_i'' \dots \dots \dots (3)$$

$$L_s = n^2 \frac{O}{l} \mu_i' \dots \dots \dots (4)$$

In these equations ω is the angular velocity of the sinusoidally alternating field.

Permeability and losses in strong fields

At high values of the flux density, the permeability and losses cannot be described by a complex permeability $\mu = \mu' - j\mu''$. In the first place, at high values, the relationship $B = \mu H$ is no longer fully valid, even if μ is assumed to be dependent on the frequency and the peak value of the flux density. This is associated with the distortion (to be discussed more fully later) whereby a field H which varies purely sinusoidally with time, produces a non-sinusoidal flux density B . In the second place, the use of a permeability μ would not result in any simplification, since this quantity would depend too closely on the frequency and the flux density. The behaviour of Ferroxcube can thus not be described in a simple manner in the case of strong fields, as is possible with weak fields. As already mentioned in the introduction, recourse must be had to graphical representations: one giving the relationship between the amplitudes B_m and H_m for various frequencies, another from which can be read the power dissipated by a sinusoidally alternating field in a Ferroxcube core, per unit volume, as a function of the maximum flux density in the core (see I, figs 7, 8 and 9), and finally a further figure giving the distortion as a function of frequency and maximum field strength.

Measurement of permeability and losses in "weak" fields

Measurements at low frequencies (50 kc/s—10 Mc/s)

The most obvious manner of determining the quantities μ_i' and μ_i'' at relatively low frequencies (e.g. up to 10 Mc/s) follows directly from the definitions of these quantities and their relationship to the quantities R_s and L_s for a coil with a Ferroxcube core as given in the introduction. A number of turns of stranded copper wire are wound on a previously demagnetized ring of square or circular cross-section, made of the material to be investigated. The inductance L_s and the series resistance R_s of the resulting coil are measured by means of a bridge circuit, from which, by use of formulae (3) and (4) — the dimensions of the ring and the number of turns being known — μ_i' and μ_i'' may

be calculated. In such a measurement it must be remembered that the stray capacitance C_p of the winding on the core interferes with the direct measurement of the values R_s and L_s . The values found are, $R_s/(1-\omega^2 L_s C_p)^2$ and $L_s/(1-\omega^2 L_s C_p)$, provided that both $\omega^2 L_s C_p$ and $\omega^2 L_s C_p \tan \delta_t$ are small compared with unity.

A method whereby the series inductance and the tangent of the loss angle may be measured directly is the following resonance method, which may be applied, in general, in the frequency range between 50 kc/s and 10 Mc/s (see fig. 1).

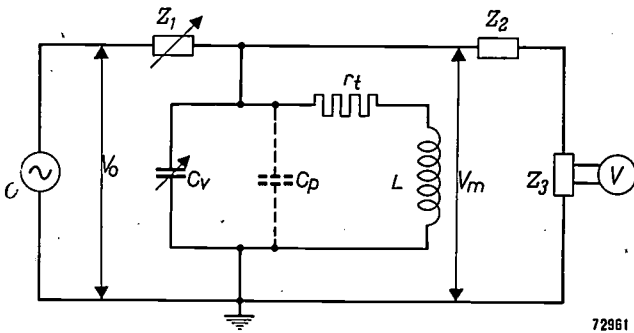


Fig. 1. Diagram of the resonance method for the determination of μ' and μ'' in "weak" fields and at "low" frequencies. O oscillator, C_v variable capacitance for the tuning of the resonant circuit, L inductance, r_t equivalent resistance for the total losses in the resonance circuit, C_p stray capacitance, V voltmeter.

An accurately calibrated oscillator (O) is coupled via a large impedance Z_1 to a tuned circuit. The output voltage V_0 of the oscillator is independent of the frequency and of the impedance of the measuring circuit. The voltage V_m across the tuned circuit is measured by an electronic voltmeter (V), which is very loosely coupled to the tuned circuit of impedance Z via the impedance Z_2 and its own input impedance Z_3 . If $Z_1 \gg Z$ and also $Z_2 \gg Z$, then the current i through the measuring circuit is, in close approximation, determined exclusively by the magnitude of Z_1 , and at the same time the voltage V_m across the tuned circuit is determined exclusively by i and the magnitude of Z .

The tuned circuit itself comprises the following elements: the toroid, with a Ferroxcube core and inductance L_s , and a resistance r_t representing all the losses in the tuned circuit, a variable capacitor with capacitance C_v serving for the tuning of the circuit, to which must be added the stray capacitance C_p of the coil with core.

The resistance r_t comprises the various components:

$$r_t = r_F + r_D + r_K, \dots \dots (5)$$

in which r_F represents the ferromagnetic losses

occurring in the core, r_D the dielectric losses in the core, and r_K the I^2R losses in the winding. If the losses are small, according to (5) the following relation also holds:

$$\tan \delta_t = r_t/\omega L_s = \tan \delta_F + \tan \delta_D + \tan \delta_K. (6)$$

The measurements are now carried out as follows: the value C_v^0 of C_v is determined, for which, at a given oscillator frequency, the voltage across the tuned circuit is at a maximum. We thus obtain C_v^0 from the condition:

$$\frac{d|Z|^2}{dC_v} = 0,$$

in which:

$$Z = \frac{(r_t + j\omega L_s)/j\omega C_t}{r_t + j\omega L_s + 1/j\omega C_t} \text{ and } C_t = C_v + C_p.$$

C_v^0 is then given by:

$$\omega^2 L_s C_t^0 = \frac{1}{1 + \tan^2 \delta_t},$$

where $C_t^0 = C_v^0 + C_p$ and $\tan \delta_t = r_t/\omega L_s$. If $\tan^2 \delta_t$ is neglected in comparison with unity, we derive the inductance desired from the resonance capacitance C_v^0 by:

$$L_s = \frac{1}{\omega^2 (C_v^0 + C_p)} \dots \dots (7)$$

We then find the loss angle δ_t by determining the difference ΔC_v between the two values of C_v for which the impedance of the tuned circuit is exactly $1/\sqrt{2}$ times the maximum value. ΔC_v is also known as the resonance width. The following now holds:

$$\tan \delta_t = \Delta C_v / 2(C_v^0 + C_p) \dots \dots (8)$$

The limitations of and sources of error in the method described are the following. In the first place $\tan \delta_t$ must always be small, as otherwise the resonance curve will be too flat, so that the resonance value C_v^0 of the measuring capacitor can no longer be determined with sufficient accuracy. In addition, the stray capacitance C_p of the winding, unless this capacitance is separately measured and allowed for in the calculations, introduces errors in the computed values of the inductance and $\tan \delta_t$; this is obvious from formula (7). From the same formula it can also be seen that the stray capacitance prevents any determination of L_s above a certain frequency, since (7) may be expressed as: $L_s \omega^2 C_t = L_s \omega^2 (C_v^0 + C_p) = 1$. Thus the value of the variable capacitance C_v at constant L_s decreases with increasing ω .

If the frequency is so high that $L_s \omega^2 C_p \approx 1$, then it is no longer possible to determine accurately the

value of C_v which is required for resonance, and at still higher frequencies resonance is not even to be attained. An attempt may be made to raise this limiting frequency by putting fewer turns on the Ferroxcube core, so that both L_s and C_p become smaller. There is a limit to what can be done in this direction, however, since by reducing the number of turns the leakage field becomes greater. This will be referred to later.

The measurement is, however, made impossible by other causes before the limiting condition $\omega^2 = 1/L_s C_p$ is reached, either because of the sharply increasing losses in the Ferroxcube ring ($\tan \delta_F$ no longer small compared with unity), or because the separate determination of the stray capacitance, the accurate knowledge of which becomes increasingly important with rising frequency — as a consequence of a decreasing C_v^0 — is no longer easily possible.

A further complication arises from the fact that the measured value of $\tan \delta_t$, in addition to the quantity $\tan \delta_F$ needed in the determination of μ_i'' , contains also $\tan \delta_D$ and $\tan \delta_K$. These (dielectric and I^2R) losses must therefore be determined separately.

Finally, a further complicating factor is the stray self-inductance of the coil, comprised by the self-inductance of the connecting leads and the self-inductance of the leakage field of the coil; the formulae (3) and (4) are only valid for a toroid which has no leakage field.

The errors introduced by the causes summarized here may be minimized in the following way.

The self-capacitance and the dielectric losses in the core may be reduced by inserting a layer of styroflex between the core and the winding (styroflex is a foil of polystyrene, a material of low dielectric constant, which introduces practically no dielectric losses). This layer prevents the lines of electric flux present between two adjacent turns from running through the Ferroxcube (see fig. 2). Because of the high values of the real (ϵ') and the imaginary part (ϵ'') of the dielectric constant of certain Ferroxcube materials (see I, fig. 10), this would of course result in both a large stray capacitance and large dielectric losses.

The I^2R losses in the copper winding may be kept small by the use of stranded wire, the thickness of the separate wires of which must be matched to the measuring frequency. The fact that there exists for every frequency a wire thickness which is preferably not exceeded, is, as is known, to be ascribed to the presence of the skin-effect. Up to about 10 Mc/s, there is a rule of thumb according

to which the thickness d , in microns, of the separate wires, must be smaller than the wavelength in metres.

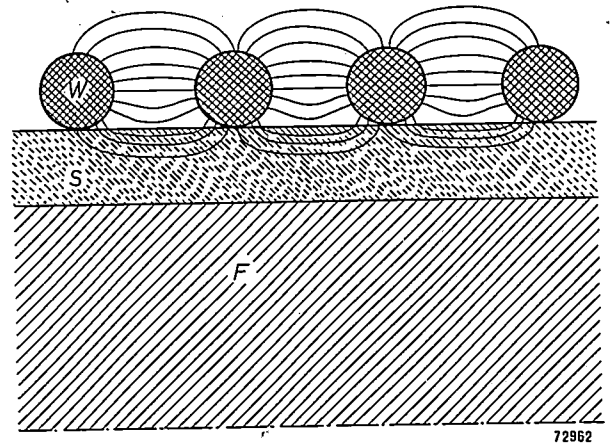


Fig. 2. Influence of a layer of styroflex between the ferromagnetic core F and the turns W . The lines of electric force between two adjacent turns run almost exclusively either in the air or in the styroflex S , which introduces only negligible losses.

Finally, the stray self-inductance of the connecting leads and of the leakage field can be reduced by using short leads lying closely adjacent to each other and by employing a large number of turns. This, however, is at the expense of the measuring range (see above), so that it is necessary to come to a compromise.

The above description of the resonance method is based on the supposition that the measurements are carried out by variation of a measuring capacitance C_v at constant frequency. It is, however, possible to use the analogous method whereby measurements are made by variations of frequency at a constant capacitance C_t . Such a method has the advantage that the stray capacitance need not be determined separately if the loss angle is to be measured, since $\tan \delta_t = \Delta f/f_0$; it can, however, lead to difficulties if the permeability and the losses in the frequency range near the measuring frequency depend sharply on the frequency.

Measurements at high frequencies (5 Mc/s-3000 Mc/s)

The method now to be described, for high frequencies (5 Mc/s-3000 Mc/s), is also a resonance method, which, however, no longer uses "lumped" circuit elements. Use is made, rather, of a resonant coaxial line³⁾. The special advantage of this method is that no turns have to be wound on the core and

³⁾ Regarding the coaxial line, see for example: C. G. A. von Lindern and G. de Vries, Philips tech. Rev. 6, 240, 1941. For a more extensive treatment see, for example, R. W. P. King, H. R. Mimmo, A. H. Wing, Transmission Lines, Antennas and Waveguides, Mc Graw Hill Book Co., New York, London, 1945.

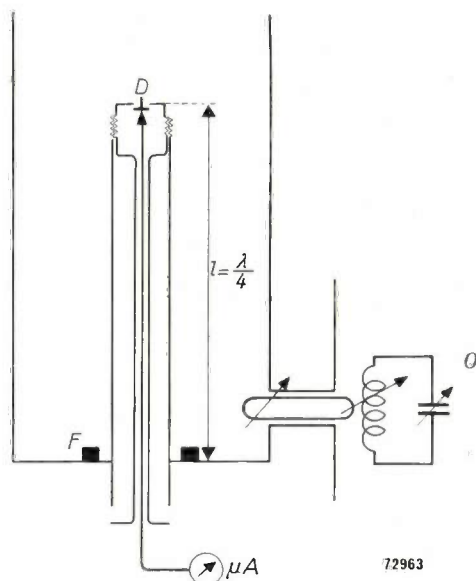


Fig. 3. Coaxial line ($l = \lambda/4$) for the measurement of μ' and μ'' at high frequencies. *O* oscillator, *D* detector, coupled to the electric field, *F* Ferroxcube ring.

that all the inherent sources of error and limitations are eliminated.

The frequency range under consideration must be further divided into two parts. A coaxial $\lambda/4$ resonator for the lowest frequency of this range would be much too long to be used in practice. It

is therefore more practical to use an open $\lambda/4$ resonator (see *figs 3 and 4*) in the frequency range of 100 Mc/s-3000 Mc/s — or, at the highest frequencies, for greater convenience, a $(2n + 1) \lambda/4$ resonator — whilst for the range of 5 Mc/s-100 Mc/s a coaxial resonator is used in combination with a “lumped” capacitance which is connected to the upper end of the line (top capacitance) (see *figs 5 and 6*); a coil may also be connected in series in the circuit. This small change in the set-up has, however, no consequences of a fundamental character.

The measuring equipment is set up as follows. A coaxial line of a certain length l is coupled to an oscillator by means of a magnetic coupling loop. The voltage existing at the upper end of the resonator is determined with the aid of a crystal detector, which is placed inside the inner conductor, at the top. For convenience of measurement it is essential that the couplings between the coaxial resonator and the oscillator and between the detector and the resonator are so weak that the amplitude and frequency of the oscillator are independent of the setting of the resonator, and also that the resonator is not influenced by the detector.

The measurement of the real and the imaginary parts of the initial permeability μ_i of Ferroxcube

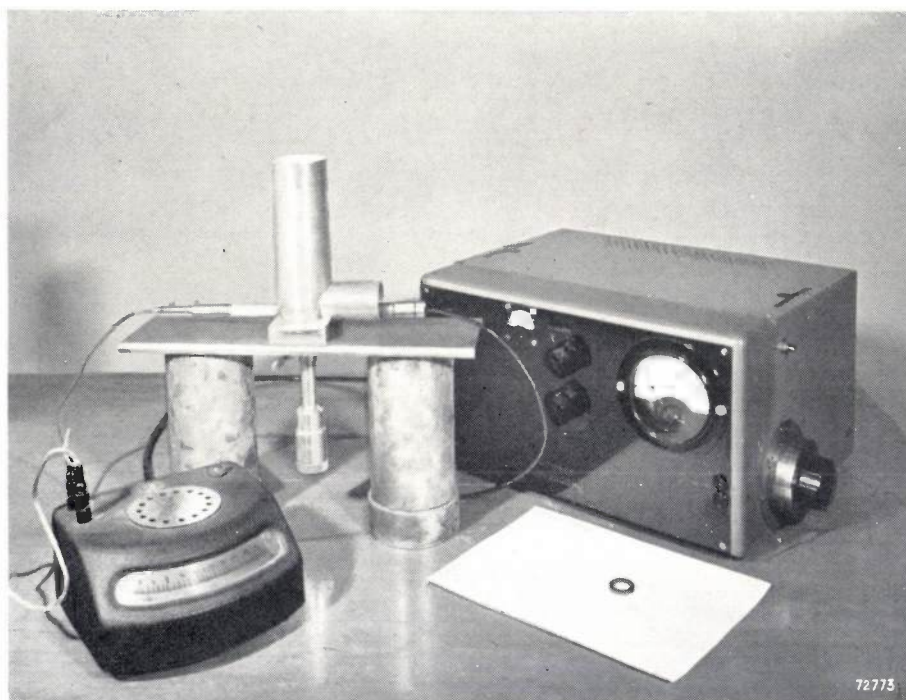


Fig. 4. Photograph of the equipment shown diagrammatically in *fig. 3*. At the right, the oscillator, in the centre, a coaxial $\lambda/4$ resonator suitable for frequencies from 500-2000 Mc/s, with adjustment screw underneath for variation of the length. The projecting portion of the resonator at the left is the detector, which, however, different from the position shown in *fig. 3*, is coupled to the magnetic field. In front of the oscillator is a Ferroxcube ring.

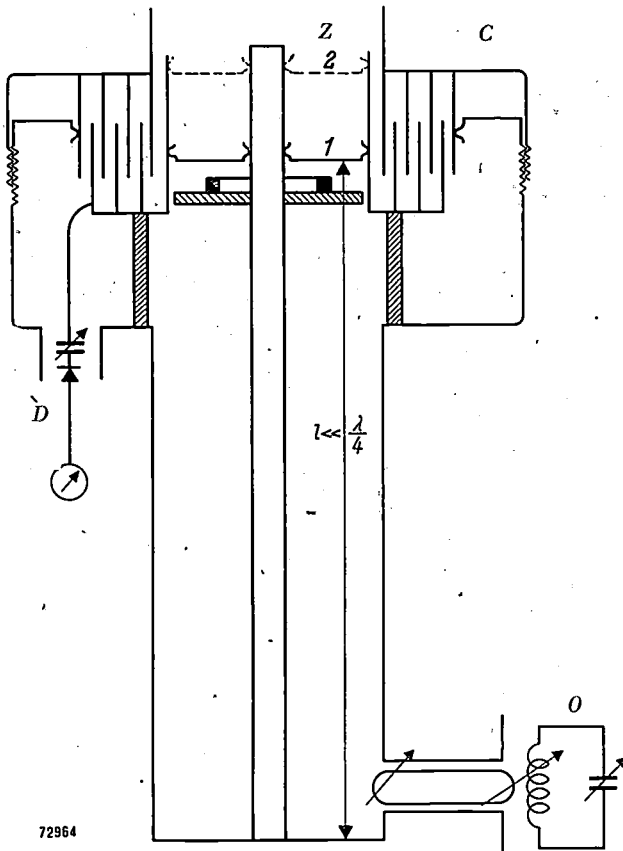


Fig. 5. Coaxial line ($l < \lambda/4$) in combination with a lumped capacitance. *O* oscillator, *D* detector, *C* lumped variable capacitance, *Z* movable piston in two positions 1 and 2, with Ferroxcube ring underneath.

at a given frequency is now carried out in a series of steps ⁴):

1a) Determination of the resonance length l^0 (or, in the case of a line terminated by a capacitor, of the resonance capacitance C^0) of the empty line.

1b) Determination of the difference Δl_0 (ΔC_0) between the two values for the empty line at which the voltage is $1/\sqrt{2}$ times its maximum value. From the resonance width thus obtained, the quality factor Q_0 of the empty line is derived: $Q_0 = l^0/\Delta l^0$ ($C^0/2\Delta C_0$).

2a) A ring of Ferroxcube material is introduced into the coaxial line at a point where the electric field strength is practically zero. The length variation δl_F (or alternatively the capacitance variation δC_F) necessary to adjust the voltage at the top end of the resonator to its maximum is now determined.
2b) Determination of the resonance width, expressed as Δl_F , or alternatively ΔC_F , for the line with Ferroxcube ring.

⁴) For the sake of simplicity the description of the case where the frequency of the oscillator is varied will be omitted. The procedure is closely analogous and affords no further insight. Further, the remarks made at the end of the previous section regarding the frequency variation method apply here also.

In principle all the data are now available which, together with the dimensions of the ring and the measuring line, are necessary for the determination of the two components of the permeability of the Ferroxcube. The formulae can be simplified, however, if one first measures for a copper ring the quantity δl_K (δC_K) corresponding to the quantity δl_F (δC_F), the copper ring having the same dimensions as the Ferroxcube ring ⁵). Because of the skin-effect the magnetic field penetrates only to a very small depth into the copper ring, so that in the frequency range employed the copper ring can, to a close approximation, be regarded as a ring with magnetic permeability $\mu' = 0$.

If the quality factor of the line with Ferroxcube ring is sufficiently high:

$$(l^0 + \delta l_F)/\Delta l_F \text{ or } (C^0 + \delta C_F)/2\Delta C_F > 10,$$

and if the relative detuning $\delta l_F/l^0$ or $\delta C_F/C^0$ is small in comparison with unity, then:

$$\mu_i'/\mu_0 - 1 = -\frac{\delta l_F}{\delta l_K}, \text{ or } \mu_i'/\mu_0 - 1 = -\frac{\delta C_F}{\delta C_K}. \quad (9)$$

The imaginary part of the permeability is given by:

$$\mu_i''/\mu_0 = \frac{\Delta l_F - \Delta l_0 (1 + \delta l_F/l^0)^2}{2\delta l_K}$$

$$\text{or } \mu_i''/\mu_0 = \frac{\Delta C_F (1 + \delta C_F/C^0)^{-2} - \Delta C_0}{2\delta C_K}. \quad (10)$$

The difficulties limiting the applicability of the resonance method as carried out according to the previous section, are obviated by this coaxial set-up.

The condition that the quality factor of the empty resonator and of the resonator with Ferroxcube ring must be high, still applies because the formulae given are only accurate to the exclusion of terms of the order of $\tan^2 \delta$ and that otherwise the resonance peak is no longer sharp enough for the accurate determination of δl_F (δC_F). This does not mean, however, that only Ferroxcube materials with a low $\tan \delta_F$ may be measured by the coaxial method. The ferromagnetic ring occupies only a small portion of the volume of the coaxial line, so that the quality factor of the line can still be high even in the presence of a ring with high losses.

No stray capacitance is present in the coaxial method, since the Ferroxcube ring is placed in the line at a point where there is no electric field. For the same reason there are no difficulties due to the introduction of dielectric losses.

⁵) H. J. Lindenhovius and J. C. van der Breggen, Philips Res. Rep. 3, 37, 1948.

The I^2R losses arising from the finite conductivity of the walls of the coaxial line are, of course, present, so that the quality factor of the empty line is not infinitely high. In contrast to the application of the resonance method described earlier for low frequen-

greater than the order of 1% are to be avoided there are a few further stipulations to be complied with to ensure accurate measurements: viz., the relative detuning $\delta l_F/l^0$ must not be too great, since terms of the order of $(\delta l_F/l^0)^2$ are neglected



Fig. 6. Photograph of the equipment shown diagrammatically in fig. 5. At the right, the oscillator, on which is lying an open coaxial $\lambda/4$ line which is used in the range of 250-1000 Mc/s. At the end of the central conductor the electrically coupled detector can be seen. Standing at the left, a coaxial line with variable top capacitance, for frequencies of 5-100 Mc/s. To the left of this line is the electrically coupled detector. The helix in the foreground is the inner conductor of a line, in which the helix serves as lumped inductance (see text).

cies, the I^2R losses in the coaxial method are easily measured and allowed for in the calculations (see formula 10). In the former method the losses in the winding without core cannot be measured in any simple manner, since the distribution of the electromagnetic field is strongly influenced by the presence of the core.

Moreover, there is no stray self-inductance in the coaxial method.

Apart from the previously mentioned condition that the quality factor of the line with Ferroxcube ring may not be smaller than 10 if errors

in the formulae. This condition can always be met either by the use of a longer line, or by introducing a "lumped" inductance in series with the measuring line.

On the other hand the losses and the detuning may not be too small to permit of their sufficiently accurate measurement. In many cases this can be arranged by using a short coaxial line in combination with a large capacitance, and, possibly, by choosing the height of the Ferroxcube ring large.

In the event that a type of Ferroxcube has to be investigated which possesses a high permeability

(e.g. $\mu_i' = 1000 \mu_0$), the following problem arises. Such a Ferroxcube ring gives rise to a detuning $\delta l_F(\delta C_R)$ 1000 times as great as that needed for a copper ring of the same dimensions. Since satisfactory accuracy of measurement cannot be achieved in this way, the Ferroxcube must in this case be compared with a copper ring of much greater dimensions than the Ferroxcube ring, so that the detuning differ by not more than e.g. a factor 10. In practice such a copper ring is replaced by a movable piston in the upper part of the coaxial line (see fig. 4). The movement of the piston is then equivalent to the insertion or removal of a large copper ring of known dimensions. From the detuning measured after removing the Ferroxcube ring by lowering the piston, the detuning which would be caused by a copper ring with the same dimensions as the Ferroxcube ring can easily be calculated, upon which formulae (9) and (10) again give the components of the permeability.

The only real difficulty which must be borne in mind in the interpretation of results obtained by the coaxial method lies in the dimensional effects occurring at these high frequencies, due to the cross-sectional dimensions of the ring being no longer small compared with the wavelength⁶⁾. As a consequence of these dimensional effects, the measured values of the permeability depend on the shape and size of the Ferroxcube ring used in the measurements, so that further corrections must be introduced to derive the material characteristics (independent of the shape) from the measured values. This results, however, in complicated calculations, and demands, moreover, a knowledge of the complex dielectric constant of the material at the frequency employed. For this reason these calculations should be obviated as far as possible by choosing the height of the ring sufficiently small.

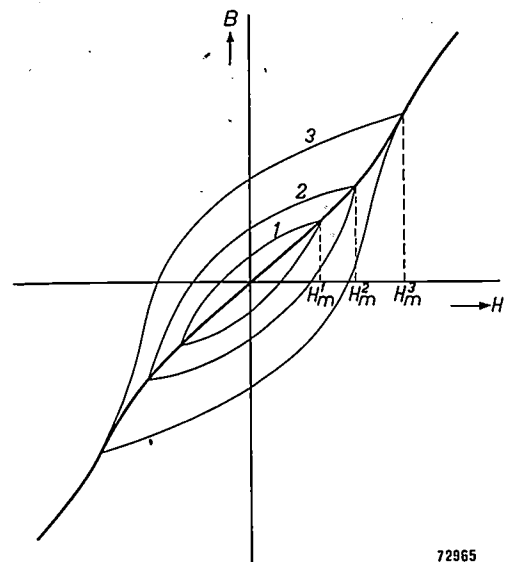
Measurements in strong fields

The magnetization curve

The magnetization curve at a given frequency implies here the relation existing between the am-

⁶⁾ See I, p. 193.

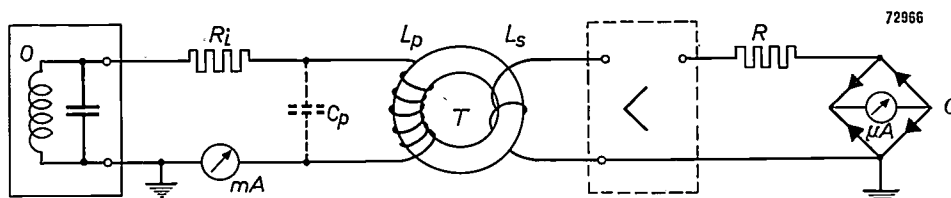
plitude H_m of a magnetic field varying sinusoidally with time and the maximum value B_m of the periodically (but in general not sinusoidally) alternating flux density in the ferromagnetic material, the latter being demagnetized before the measurement (see fig. 7). We thus deviate slightly from the accepted terminology. The definition given above is only of practical significance for ferromagnetic materials with a high resistivity, since in the case of metallic ferromagnetic materials, the magnetization curve defined in this way would depend on the shape and dimensions of the material being examined.



72965

Fig. 7. At low frequencies there is a hysteresis loop for each value of H_m . Three examples are given in the figure. The extremities of the hysteresis loops lie on the heavily drawn line, representing the magnetization curve at high frequencies.

The measurement of the magnetization curve is carried out, schematically, as follows (see fig. 8). The primary coil (inductance L_p) of a transformer with a toroidal Ferroxcube core is fed from an oscillator with variable frequency. The primary current (r.m.s. value i_p) supplied by the oscillator and measured by a milliammeter must be as purely sinusoidal as possible. The magnetic field



72966

Fig. 8. Diagram of the method for the determination of the magnetization curve. *O* oscillator, R_i resistance of the primary circuit, L_p primary inductance, C_p stray capacitance. *T*: transformer with toroidal Ferroxcube core, L_s secondary inductance, R resistance of secondary circuit, *G* full-wave rectifier circuit.

excited in the core by the primary current has a peak value:

$$H_m = \frac{i_p n_p \sqrt{2}}{l} \dots \dots \dots (11)$$

In the above relation n_p is the number of turns in the primary coil and l is the length of the centre line of the core.

The measurement of the secondary voltage of the transformer is carried out by means of a full-wave rectifier circuit as shown in fig. 8. As a consequence of the full-wave rectification the reading of the microammeter is proportional to the average value of the secondary voltage $E_s(t)$ over a half-cycle, with the initial condition $E_s(0) = 0$:

$$\bar{E} = \frac{\omega}{\pi} \int_0^{\frac{\pi}{\omega}} E_s(t) dt. \dots \dots \dots (12)$$

$E_s(t)$ is related to the derivative of the magnetic flux Φ as follows:

$$E_s(t) = -\frac{d\Phi}{dt} = -n_s O \frac{dB}{dt},$$

in which n_s is the number of secondary turns and O the cross-sectional area of the core. By substitution in (12) we find:

$$\bar{E} = -\frac{\omega}{\pi} n_s O \left[B\left(\frac{\pi}{\omega}\right) - B(0) \right] = 2 \frac{\omega}{\pi} n_s O B_m,$$

and thus:

$$B_m = \frac{\pi \bar{E}}{2\omega n_s O} \dots \dots \dots (13)$$

The limitations of and sources of error in the method described are as follows.

The value read from the milliammeter (see fig. 8) is not equal to the r.m.s. value of the primary current i_p , since there is always a stray capacitance present in parallel with the primary winding. If we stipulate that the current flowing through the stray capacitance is less than 1% of the current flowing through L_p , the following condition must be satisfied:

$$\omega^2 L_p C_p \leq 10^{-2} \dots \dots \dots (14)$$

Owing to the non-linear relationship between B and H at the field strength under consideration, the flux in the core will no longer be purely sinusoidal, but will contain, inter alia, a component varying with time in proportion to $\sin 3\omega t$ (third harmonic). Thus the voltage across the primary inductance will contain, apart from a component V_{p1} , which varies with the angular velocity ω , a further component V_{p3}

which varies with the angular velocity 3ω . To keep the primary current approximately sinusoidal, we must stipulate that the value of the current i_{p3} caused by the above-mentioned V_{p3} is smaller than $0.01 i_{p1}$, i_{p1} being the desired current component alternating with angular velocity ω . It can be assumed that for most cases the e.m.f. due to the third harmonic is smaller than $0.1 V_{p1} = 0.1 i_{p1} \omega L_{p1}$. From this it follows that i_{p3} is smaller than

$$\left| \frac{0.1 i_{p1} \omega L_{p1}}{R_i + j 3\omega L_{p3}} \right|,$$

and, according to our stipulation, this must also be smaller than $0.01 i_{p1}$. In this expression R_i is the resistance in the circuit in series with the primary coil, which is practically the same as the total external impedance for the third harmonic. If this condition is further developed, the conclusion is reached that, if L_{p3} is assumed not to differ sharply from L_{p1} , then:

$$R_i \geq 9\omega L_p \dots \dots \dots (15)$$

The branch containing the stray capacitance C_p must also, of course, have a sufficiently large impedance for i_{p3} . From this arises the condition:

$$13 \times 3\omega L_p \leq \frac{1}{3\omega C_p},$$

which is practically the same as the condition (14).

The maximum field strength which can be employed for measurements is limited both by the conditions (11) and (14), and by the maximum current value which the oscillator can supply. Thus, for a given frequency $\omega/2\pi$ and a given stray capacitance C_p , L_p and hence the number of turns n_p is limited.

At high frequencies and high field strengths difficulties may also be experienced by the large amount of heat developed in the core. This latter, coupled with the low thermal conductivity of the Ferroxcube material, makes it essential that all measurements are carried out quickly and that a long period must elapse between two measurements.

Losses at high values of flux density

As was amply shown earlier in I, various physical processes contribute to the losses in the range of high values of the flux density. These various contributions cannot easily be separated in a general way. In this range of high flux densities it is therefore only possible to investigate the magnitude of the total losses occurring in a unit volume of a Ferroxcube

core when a purely sinusoidal, alternating magnetic field of large amplitude is present.

There are two different methods available for determining this quantity. In the first method the heat developed in the Ferroxcube core by the alternating electromagnetic field is measured directly in a calorimeter. The second method is similar to the coaxial method described in a previous section.

a) Calorimetric method

In the calorimetric method a winding on a ring-shaped core of Ferroxcube is again employed. The coil with core is placed in a calorimeter, and the coil is supplied with a sinusoidal current. The total quantity of heat developed in a certain time due to the energy dissipation in the core can now be determined from the temperature rise in the calorimeter. This method offers the advantage that one can in this way determine with certainty *all* the losses arising in the core due to the electromagnetic field. One measures thus not only all types of ferromagnetic losses but also the dielectric losses. This appears to be a useful method for flux densities exceeding 50×10^{-4} Wb/m² and in a frequency range between 50 kc/s and 1 Mc/s. The stipulation that the frequency may not be too low and the flux density not too small is occasioned by the sensitivity of the calorimeter; the heat developed per unit time may not be too small. That the calorimetric method is not suitable for frequencies over a certain value is a consequence of the self-capacitance C_p of the winding on the core, mentioned repeatedly above. This results in the measured value of the current through the winding not being the true value, so that an exact determination of the field strength is no longer possible.

Apart from these limitations, a further disadvantage of the calorimetric method is that the measurements require much time. This method is thus not suitable for routine investigations.

b) The coaxial method

The limitations of the calorimetric method are avoided in the coaxial method, although this in turn is subject to other limitations. The latter method can be satisfactorily applied for flux densities lying below 100×10^{-4} Wb/m². Hence measurements can also be carried out in the range not encompassed by the calorimetric method (below 50×10^{-4} Wb/m²). The frequency range of the measuring apparatus made in the Philips laboratory in Eindhoven lies between 50 kc/s and 10 Mc/s.

As we have already mentioned, the coaxial method described here is similar to that already

discussed on pages 248-252, and which was intended for use at very small values of flux density. For measurements at high values of flux density, however, the power of the oscillator must be considerably greater. Because of the low frequency, the measuring circuit always contains a "lumped" inductance and capacitance.

As with the coaxial method for measurements at very low values of flux density, the change in the resonance capacitance and the change in the width of the resonance curve due to the insertion of the Ferroxcube ring into the resonator, are determined in this case also. This amounts to the determination of the real (μ') and the imaginary (μ'') components of the magnetic permeability (μ). However, in the introduction it was shown that the values μ' and μ'' become meaningless at high values of flux density. The contradiction between these two statements is only apparent, as we shall now see. The two measured quantities μ' and μ'' are no longer the real and the complex components of a permeability defined by a relation $B = \mu H$. Rather must this relation be replaced by one of the form:

$$B = (\mu_i + \nu H_m) H \pm \nu (H_m^2 - H^2), \quad (16)$$

in which H_m is the maximum field strength. The + and - signs must be used for the upper and lower branches of the hysteresis loop respectively; ν is the Rayleigh constant. By the coaxial method, therefore, one thus measures the components of the first term of the right-hand side of (16). The error involved is not easily estimated. The complexity and confusion of this situation (and the resulting limited accuracy of the results) is one of the disadvantages of this method. In general, one can say that not the total losses arising in the Ferroxcube ring are determined, but principally the losses corresponding to that part of the magnetic flux density having the same periodicity as the magnetic field. This is associated with the fact that the peak value of the voltage across the resonator is measured by this method. This value is mainly dependent on the flux density component with the fundamental frequency. The voltages due to the higher harmonics, which add with an unknown phase to the voltage of the fundamental frequency, are small since there is no resonance for the higher harmonics.

For flux densities up to about 50×10^{-4} Wb/m² this has no unfavourable consequences of any importance in practice: comparison with loss measurements by the calorimetric method shows that errors made with the coaxial method in the range of the

flux density mentioned are in general small. If, however, reliable results in the range above $50 \cdot 10^{-4}$ Wb/m² are desired, then resort must be taken to the elaborate calorimetric method.

Distorsion

From the above-mentioned relation (16) between B and H it follows that, at large values of the maximum field strength H_m , deviations from the linear relationship between current and voltage appear. If, thus, a purely sinusoidal current of angular velocity ω flows through the primary winding of a transformer, a voltage is produced across the secondary winding which may be described by the following series ⁷⁾:

$$V = V_1 \sin \omega t + V_3 \sin 3\omega t + V_5 \sin 5\omega t + \dots \quad (17)$$

The distortion is now defined by:

$$D = \sqrt{\frac{V_3^2 + V_5^2 + \dots}{V_1^2}}$$

In general, however, the higher harmonics V_5 etc. may be neglected, so that by approximation:

$$D \approx \frac{V_3}{V_1} \dots \dots \dots (18)$$

The most direct method of measuring the distortion is the following. The voltage across the secondary winding of a transformer is measured by means of a selective voltmeter. In this way the values V_1 and V_3 are at once determined.

More usual, but, as we shall see, not to be used without caution, is another procedure which we shall describe more fully.

For sufficiently weak fields there is a simple relation between the distortion V_3/V_1 , the angular velocity ω of the fundamental wave, the inductance L of the (primary) coil and the hysteresis resistance R_H ⁸⁾:

$$\frac{V_3}{V_1} = \frac{0.6 R_H}{\omega L} \dots \dots \dots (19)$$

The existence of this relationship introduces the possibility of determining the distortion by measurement of the hysteresis resistance.

The hysteresis resistance, which represents the loss caused by hysteresis in a coil with a ferromagnetic core, is a quantity depending on the maximum field strength. If we assume that the loss due to hysteresis is the only type of loss depending on

the field strength, which in general is certainly not the case ⁹⁾, then the hysteresis resistance R_H may be found from a graph in which the total resistance R_t is plotted as a function of the maximum field strength H_m according to the relation:

$$R_H = R_t - R_0, \dots \dots \dots (20)$$

in which R_0 is the part of the resistance independent of the field strength. At "low" frequencies R_t may be measured, for example, with the aid of a Maxwell bridge ¹⁰⁾.

With the same conditions as for (19), it can be deduced that, for a given value of H_m , R_H increases proportionally to the frequency. Measurements on Ferroxcube materials have shown, however, that above a certain frequency range (depending on the material) the value of the hysteresis resistance found according to (20) is greater and increases more sharply with frequency than is to be expected from the results of distortion measurements with the aid of a selective voltmeter. This indicates that at high frequencies another loss mechanism occurs which also depends on the maximum field strength, so that the relation (20) can no longer be used to determine, from the measured value R_t , the resistance R_H , which is associated only with hysteresis phenomena. This is clearly seen in fig. 9.

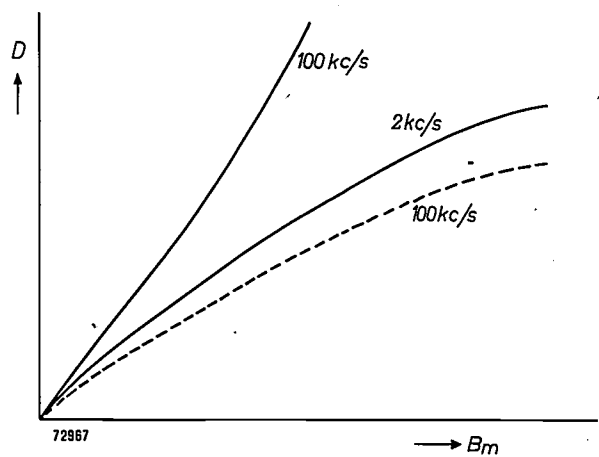


Fig. 9. Distortion D in Ferroxcube as a function of the maximum flux density B_m , at 2 kc/s and at 100 kc/s (schematic). The full line for 100 kc/s gives the distortion as would be derived from the apparent hysteresis resistance. The broken line gives the results of measurements with a selective voltmeter. At 2 kc/s the results by both methods are identical.

This gives the distortion schematically as a function of the maximum flux density for two different frequencies. The full line for 100 kc/s gives approximately the distortion to be found by calculation

⁷⁾ The fact that no even harmonics appear in this series is due to the symmetry of the hysteresis loop about the origin of the co-ordinates for B and H .
⁸⁾ E. Petersen, Bell Syst. tech. J. 7, 762, 1928.

⁹⁾ For a full discussion of this hypothesis see I.
¹⁰⁾ See for example D. Polder, Proc. I.E.E. 97, 246, 1950.

from the "apparent" hysteresis resistance according to formulae (20) and (19); the broken line for 100 kc/s gives approximately the values found by the method using the selective voltmeter. Whilst the results found by the two methods agree for very low frequencies (see the curve for 2 kc/s), one sees that at the higher frequencies a marked discrepancy appears. The distortion can thus not be determined at high frequencies by use of the relations (20) and (19), and recourse has to be made to the rather less simple method already mentioned, using a selective voltmeter.

Summary. For the measurement of the permeability and the losses of Ferroxcube materials in weak fields, a resonance method is applied in which the measuring circuit consists of either lumped network elements (for a frequency range of

about 50 kc/s to 10 Mc/s), or of a coaxial line with or without a lumped inductance or capacitance (for frequencies in the range between 5 Mc/s and 3000 Mc/s). The sources of error and limitations of both methods are amply described. In strong fields a determination of the magnetization curve replaces the permeability measurement. This is carried out with the aid of a full-wave rectifier circuit in which the secondary voltage of a transformer with Ferroxcube core is measured. Two methods are available for the measurement of losses in strong fields. In the first method the quantity of heat generated in a Ferroxcube ring placed in a sinusoidally alternating magnetic field is measured in a calorimeter. This method is suitable for flux densities above 50×10^{-4} Wb/m² and frequencies below 1 Mc/s. A disadvantage is that a measurement is rather elaborate. For smaller flux densities and at higher frequencies use can again be made of a coaxial line. As a result of the distortion inherent in this method, however, not the total losses but only those caused by the fundamental frequency are measured. This is no disadvantage below 50×10^{-4} Wb/m². Finally, some remarks are made on the measurement of the distortion. At very low frequencies the distortion may be determined from measurements of that part of the resistance which depends on the maximum field strength, whereas at high frequencies the distortion must be measured directly with the aid of a selective voltmeter.

A DIRECT READING PRECISION POLAROGRAPH

by H. A. DELL *) and C. H. R. GENTRY **).

545.33:621.317.755

For the last few years it has been usual in metallurgical investigations to measure the concentration of certain metal ions with the aid of a polarograph. An objection against the use of this instrument is that the taking of a polarogram and the deriving of the desired results take up much time. The polarograph described in the following removes this objection as the results can be read directly from the instrument.

Introduction

There are various types of electrical measuring apparatus in use in chemical analysis at the present time. An example is the polarograph, which is used extensively in metallurgical investigations. With this instrument the current which flows when a particular voltage is applied between two electrodes located in the liquid to be analysed can be measured. From the form of the curve giving the current as a function of voltage, the so-called polarogram, the presence of certain metal ions in a solution can be investigated qualitatively and quantitatively.

The anode may be a mercury pool, a calomel or other electrode, non-polarizable under the conditions to be studied. The cathode consists of a small drop of mercury which is suspended from a capillary tube.

As the currents to be measured are of the order of a few microamperes, a galvanometer is usually employed, making the apparatus rather delicate. As far back as 1939, a description was given in this periodical of a method whereby the galvanometer was replaced by a cathode ray oscillograph or a measuring bridge, giving not only a more robust set-up but also an appreciable shortening in the time required for the analysis of a given liquid ¹⁾. For a qualitative analysis it is in this case no longer necessary to take a full polarogram and to measure it afterwards; this is replaced by an indication on the cathode ray tube and a reading of a calibrated scale.

In the following we shall give a description of an instrument which likewise uses a cathode ray tube as indicator, but which makes possible not only qualitative but also quantitative analysis. With the polarographs usually employed conclusions about the composition of the liquid to be analysed can only be drawn after plotting a complete polaro-

gram. With the instrument to be described, however, the concentration of certain metal ions can be deduced from the difference of two readings, resulting in a considerable saving of time. Moreover, the measurements are as exact as or even more exact than is usually the case in polarography ²⁾.

The polarograph in question is shown in *fig. 1*. Before describing this apparatus, we shall first recapitulate a few particulars concerning electrolysis with the mercury-drop electrode. (See also the article referred to in footnote ¹⁾).

Electrolysis with a mercury-drop electrode

If one applies, between two electrodes immersed in a metallic salt solution, a D.C. voltage which is gradually increased from zero, no appreciable immediate flow of current is, as a rule, to be observed. This is due to the fact that a metal ion requires a certain voltage to discharge it. This voltage differs for ions of different types. Only at a voltage that is sufficiently high can an ion of the element which has the lowest deposition voltage join with an electron from the cathode to form a metal atom. If this voltage is reached the electrons which are drawn from the cathode are replenished by the voltage source, so that an electric current is produced. If the solution is not too concentrated, and if a cathode of small surface area is used, then with a further increase of voltage the position is soon reached at which every metal ion arriving at the cathode is immediately discharged. The current then depends on the ion concentration, the surface area of the cathode and the speed with which the ions migrate to the cathode. The speed of migration of the ions considered is independent of the voltage, provided that there exists in the solution

*) Mullard Research Laboratory, Salfords, England.

***) Mullard Material Research Laboratory, Mitcham, England.

¹⁾ J. Boeke and H. van Suchtelen, Rapid chemical analysis with the mercury dropping electrode and an oscillograph or measuring bridge as indicator, Philips tech. Rev. 4, 231-236, 1939.

²⁾ A description of the instrument has already been given in two earlier publications: C. H. R. Gentry and D. Newson, *Metallurgia* 41, 107-111, 1949 (Dec.); H. A. Dell and C. H. R. Gentry, *Electronic Eng.* 28, 19-21, 1952 (Jan.).

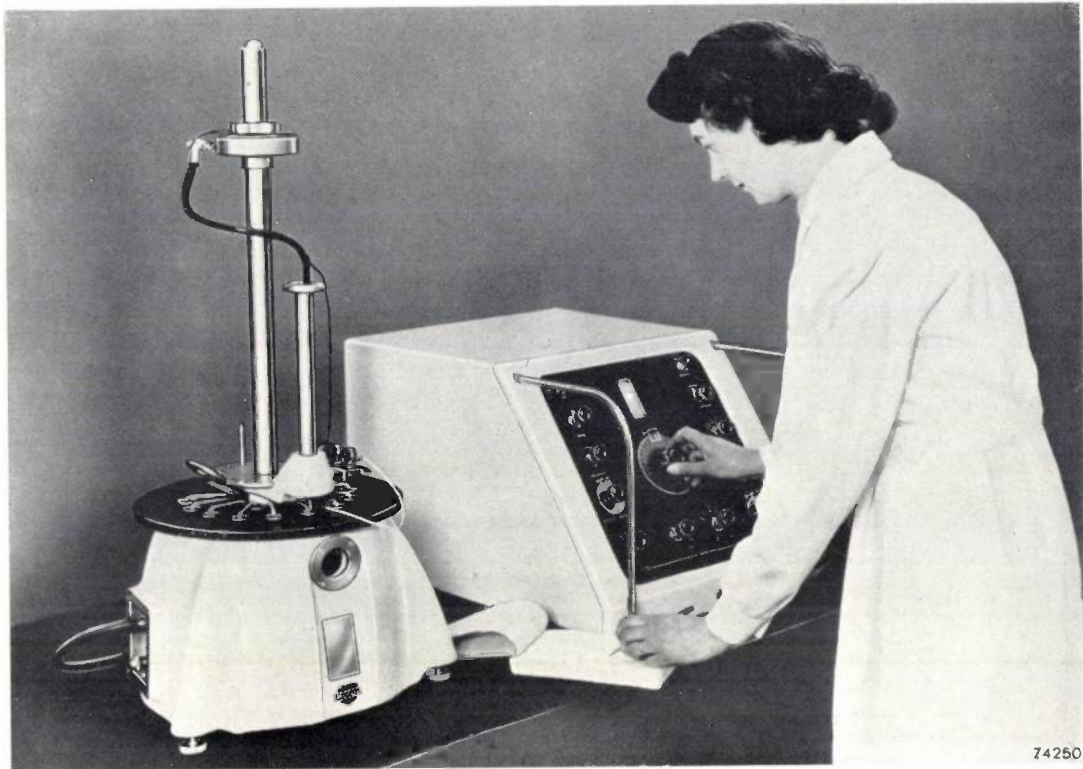


Fig. 1. The direct reading precision polarograph. On the left is a thermostat in which is placed the electrolytic cell with the solution under investigation and the mercury drop electrode.

an excess of ions having a higher deposition voltage. If the current is now plotted as a function of the voltage, a curve (polarogram) is obtained similar to that shown in *fig. 2*.

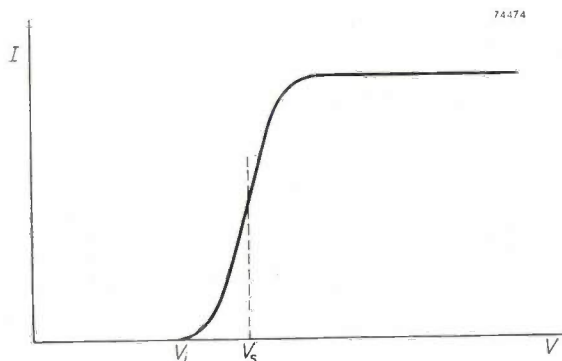


Fig. 2. Simplified representation of part of a polarogram. (The charging current of the drops is not shown.) V_i is the potential at which a current I just begins to flow in the solution, and V_s the half-wave potential.

The half-wave potential V_s is characteristic for the kind of ions considered. If various types of ions are present in the solution, current steps appear at different values of voltage. A polarogram will thus take the shape of *fig. 3*. As the magnitude of the current steps is proportional to the concen-

tration of the various ions, one can determine the latter quantity by measuring the former.

By making the cathode very small a danger is introduced that the surface will become contaminated quickly by deposition of metal. This difficulty is overcome by using as cathode a small drop of mercury hanging from a capillary tube. By supplying mercury to the tube, the drop is caused to fall off periodically, at intervals of a few seconds, thus regularly renewing the cathode surface. The falling and reforming of the mercury droplets, however, render the measurement more difficult. As the surface area of the cathode continually alters, the

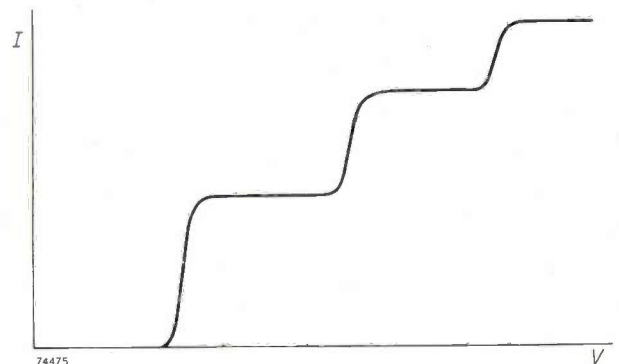


Fig. 3. Shape of a polarogram in the presence of various types of metal ions in the solution being examined.

current also changes during the life of each drop, so that for a given voltage the current varies as a function of time as shown in *fig. 4*. When the drop is very small the current is small, whilst as the drop increases in size so also does the current increase, as in curve *AB*. When the drop falls the current drops suddenly as shown by the line *BC*. It can be shown that the current during the life of each drop is proportional to the one-sixth power of the time.

In conventional polarographs a galvanometer is used with a very high inertia in order to measure the average value of the current. Since, however, the drops fall at intervals of 3 to 4 seconds, it is not possible completely to damp the oscillations of the galvanometer; to achieve this the damping would have to be so great that the form of the resulting curve would be influenced. It is therefore better only to measure the current at a certain

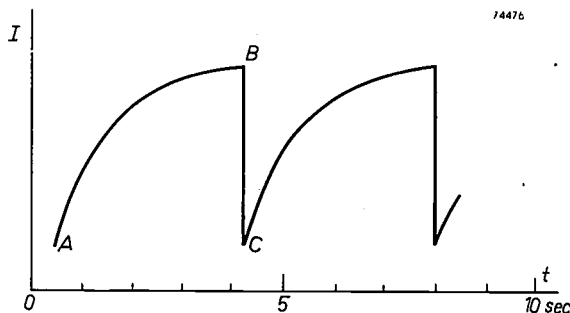


Fig. 4. Current variation in the electrolytic cell as a function of time, at constant voltage. As the mercury drop increases in size, the current rises as shown by the curve *AB*; when the drop falls, the current drops as shown by the line *BC*.

moment during the life of each drop. The most suitable time is the moment just prior to the fall of the drop, for the following reasons:

1. When a drop is detached from the cathode, the equilibrium conditions in the neighbourhood are violently upset and it is thus very difficult to interpret precisely the current early in the life of the drops.

2. The current changes most slowly at the end of the life of the drops, and consequently it is easiest to measure the current at this moment.

In the instrument to be described, therefore, the current is measured at the moments immediately prior to the falling off of the drops.

Use of an oscillograph for measurements with a polarograph

It is desirable to make up the circuit in such a way that the oscillograph is used as a zero voltage indicator, making a calibration of the instrument unnecessary. A circuit which can be used for this

purpose is shown in *fig. 5*. By means of the variable resistor R_2 a definite voltage can be applied across the potentiometer R_1 , the voltage V_1 thereafter being read on a calibrated scale on R_1 . (The resistance of

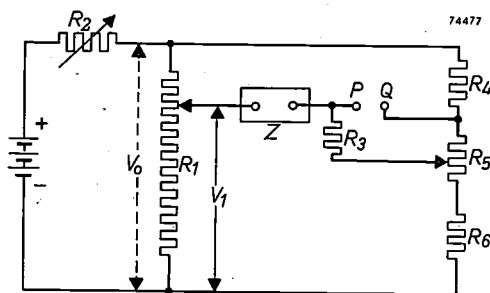


Fig. 5. A circuit whereby the current in the electrolytic cell *Z* can be measured by means of a zero voltage method. If no voltage exists between *P* and *Q*, the current in *Z* may be read from a calibrated scale on R_5 .

R_1 must be so small that the influence of the current in the electrolytic cell *Z* on the voltage distribution along R_1 can be neglected.) A resistor R_3 is connected in series with *Z*. This resistor is connected to the sliding contact of a potentiometer R_5 , which in series with the resistors R_4 and R_6 is connected across the same voltage as R_1 . The value of R_3 must be large compared with R_5 and R_6 , so that the influence of the current in R_3 on the voltage distribution across R_5 becomes negligible. If now a voltmeter is connected between the points *P* and *Q*, and the slider of R_5 so adjusted that no voltage appears between these points, then the voltage across *Z* is equal to V_1 less a fixed amount, namely the voltage on $R_5 + R_6$. During the life of a drop the current in *Z* and R_3 changes, and this is also the case with the voltage between *P* and *Q*. At the moment that no voltage exists between these points the current in *Z* and R_3 is equal to the voltage across the upper part of R_5 divided by R_3 . It is therefore possible to provide R_5 with a calibrated current scale which is valid when no voltage is present between *P* and *Q*. In this way the current in *Z* is indeed measured on a zero voltage basis.

In order to employ a cathode ray oscillograph as a zero voltage indicator, the points *P* and *Q* are connected to the vertical deflecting plates via separate two-stage amplifiers. In the lead from *P* to one of these amplifiers a resistor R_7 is introduced (*fig. 6*), and a high speed relay *X* which is connected to an alternating voltage of 50 c/s connects and disconnects the input terminals of the amplifiers 100 times per second. When now a voltage of saw-tooth waveform is applied to the horizontal deflecting plates of the cathode ray tube, a figure such as that

shown in *fig. 7* appears on the screen. The straight line *p* corresponds to zero voltage and the curve *q* represents the voltage between *P* and *Q* as a function of time.

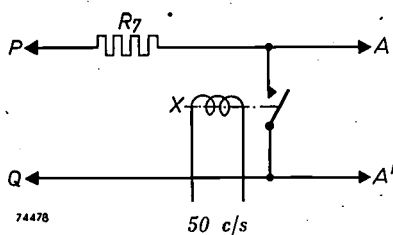


Fig. 6. Circuit permitting the use of a cathode ray tube as zero voltage indicator between points *P* and *Q* (*fig. 5*). Each of the points *A* and *A'* is connected to one of the vertical deflecting plates of the cathode ray tube via an amplifier. By means of a high speed relay *X*, the points *A* and *A'* are connected and disconnected 100 times per second.

Thus, during measurement, the potentiometer R_5 is so adjusted that the upper tips of the curve *q* just coincide with the line *p*. In this case the voltage between *P* and *Q* is zero at the moments just before the detachment of the drops. Thus the current which flows at these moments in the cell *Z* can be read on R_5 . This operation is rendered difficult, however, by the fact that the time interval at which the mercury drops fall is quite large, i.e. 1-4 seconds. The adjustment may be facilitated in two ways. Firstly use of a timebase may be dispensed with, so that two dots appear on the screen, one stationary and the other moving up and down. The slider of the potentiometer R_5 is then so adjusted that the moving point just reaches the stationary one.

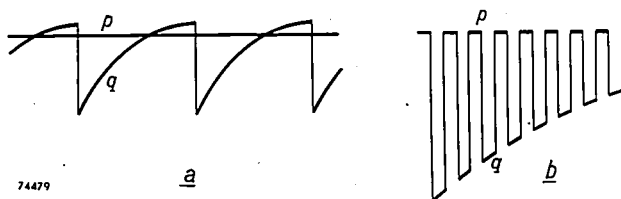


Fig. 7. a) Figure produced on the screen of the cathode ray tube by the use of the circuits of *figs. 5* and *6*. The line *p* corresponds to zero voltage and the curve *q* represents the voltage between *P* and *Q* as a function of time. b) Enlargement of part of (a). The fluorescent spot traces portions of the lines *p* and *q* alternately.

According to another method the timebase may be synchronised with the falling off of the drops, so that a standing figure preferably consisting of only one period of the current curve appears on the screen (*fig. 8*). To achieve this synchronisation, use may be made of the fact that at the moment of detachment of each drop a voltage pulse of about $\frac{1}{4}V$ appears across the resistor R_3 . It now becomes

simple, in practice, to adjust the top of the curve *q* to coincide with the straight line *p*, particularly when use is made of a tube with a long persistence screen.

The point or the line representing zero input voltage must be very stable to permit of an accurate reading. In order to make the line *p* (*figs. 7* and *8*) practically rectilinear, or the null-point stationary, it is necessary to use coupling elements, with a very large time constant in the amplifier. As a result difficulties might arise due to the long recovery time after an accidental overload. To avoid this disadvantage, use is made of D.C. amplifiers in the present instrument.

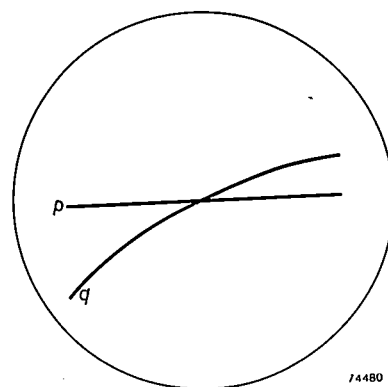


Fig. 8. By synchronisation of the time base with the growth and falling off of the drops, it can be arranged that only one period of the curve *q* appears on the cathode ray tube screen.

The tangent slope control

As already mentioned, for a quantitative measurement not only the half-wave voltages must be measured but also the magnitude of the current steps. If a polarogram were to have the form shown in *figs. 2* and *3*, this would be very simple. One would only have to measure the current at two voltages, the one a little lower and the other a little higher than the half-wave potential. In actual fact, however, the form of a polarogram differs from that shown in *figures 2* and *3* in that the flat portions have in reality a slight slope. Thus, in the neighbourhood of a current step, a polarogram takes the form shown in *fig. 9*. The reason for this lies in the fact that each newly formed droplet becomes charged, the charging current being about proportional to the applied voltage. If one now measures the current at either side of the current step, e.g. at voltages V' and V'' , then a value will be found for the current step as represented by I_1 in *fig. 9*. Actually, the size of the current step which is truly proportional to the concentration of the metal ions concerned is equal to I_2 . One can of course determine this value

by taking a full polarogram and drawing in the tangents AB and CD . This method which is in fact usually employed, is naturally a very time-consuming procedure. For this reason the circuitry

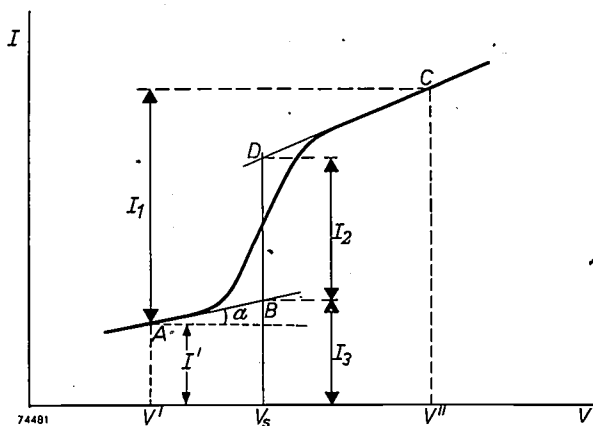


Fig. 9. Due to the charging current of the drops the straight portions of a polarogram show a certain slope. By the use of the polarograph described in this article the current step I_2 can be determined from the difference between two readings I_3 and $I_2 + I_3$.

of the polarograph described here has been extended in such a way that this graphical construction can be simulated electrically. In this way a "direct reading" instrument is obtained.

The modified circuit in which this refinement is introduced is given in fig. 10. In this arrangement two variable resistors R_7 and R_8 are connected in series with the potentiometer R_1 , the sliders of these two new resistors being coupled in such a

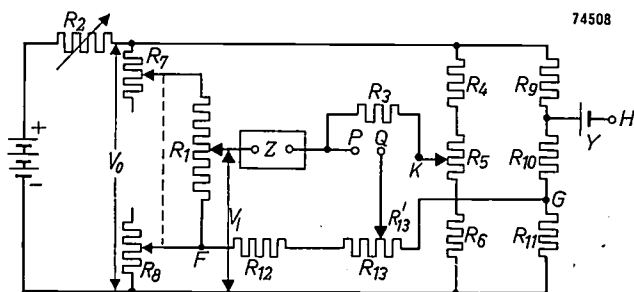


Fig. 10. Polarograph circuit by means of which the charging current of the drops can be compensated. Z electrolytic cell, Y standard cell with the help of which the voltage V_0 may be accurately adjusted to a given fixed value.

way that the total resistance $R_7 + R_1 + R_8$ remains constant on variation of R_7, R_8 . In addition a further potentiometer is introduced, comprising the resistors $R_9, R_{10},$ and R_{11} . The value of the resistors

is so chosen that the voltage at point F is the same as that at point G , when $R_7 = R_8$. An extra potential divider is connected between these points in the shape of two resistors R_{12} and R_{13} . These latter are of a sufficiently high value to ensure that their influence on the voltage distribution along R_9, R_{10} and R_{11} and along $R_7, R_1,$ and R_8 can be neglected. The point shown in fig. 5 as Q is now connected to the slider of R_{13} .

When this slider is moved to the extreme right the voltage of Q does not change if V_1 is altered by variation of R_7, R_8 . The calibration of R_5 has been carried out at this position of R_{13} , so that the current read on R_5 corresponds to the current in the electrolytic cell Z (provided the latter potentiometer is adjusted on the position where no voltage difference exists between P and Q). If $R_7 = R_8$, so that, as already mentioned, the voltage between the points F and G is zero, the moving to the left of the slider of R_{13} has no influence on this situation, since the voltage on Q does not change in this case. If however $R_7 \neq R_8$, so that the voltage of F is not equal to the voltage of G , the position of R_5 at which the voltage between P and Q is zero just before the detachment of the drops, depends on the position of R_{13} .

The procedure in measuring a current step is now as follows: After setting the slider of R_1 in the position where the voltage on the cell Z is equal to the half-wave potential V_s of the element to be determined (keeping $R_7 = R_8$), a change of voltage is introduced by means of R_7, R_8 , the voltage on Z becoming for example, equal to V' . R_5 and R_{13} are now adjusted so that at the moment before detachment no voltage appears between P and Q , the setting of these potentiometers being so chosen that for small variations of R_7, R_8 the potential between P and Q remains zero. After a few trials one soon succeeds in this. In this case the variations of the voltages at the points P and Q at small changes of R_7, R_8 are equal. The current which is now read on the calibrated scale on R_5 is not equal to the real value of the current (the calibration having been carried out with the slider of R_{13} moved completely to the right). It can be shown (see appendix) that at the positions of R_5 and R_{13} mentioned, the current read on R_5 is equal to the value which is indicated by I_3 in fig. 9.

When, now, the voltage on the cell is adjusted by means of R_7, R_8 to a value which lies on the other side of V_s , e.g. V'' , the value $I_2 + I_3$ can be measured in the same way. The difference of these readings yields the magnitude of the current step I_2 .

The complete apparatus

In the final form of the instrument, the resistor R_3 can be changed in five steps. Currents up to 100 microamperes can be measured to an accuracy of 0.1 microampere, differences of 0.01 microampere being still visible. The resistors are so chosen that the voltage applied to the electrolytic cell can be varied between 3.0 V in one direction and 0.3 V in the other direction, which is quite adequate for nearly all ion concentrations to be measured.

For the accurate adjustment of the voltage V_0 to the value at which R_1 and R_5 are calibrated, the voltage across R_{10} is compared with that of a standard cell Y . This may be done in a simple way with the oscilloscope which is already present. For this purpose the amplifiers which during the measurement connect the points P and Q to the deflection plates of the cathode ray tube can be connected to the points which are indicated by H and G in fig. 10.

It appears to be very important to keep the temperature of the solution to be investigated constant. For that reason the electrolytic cell is placed in a thermostat, keeping the temperature constant within 0.05 °C.

The accuracy with which the concentration of certain metal ions can be measured with the instrument described amounts to 0.3 per cent, which compares favourably with the 1 per cent obtainable with existing commercial instruments.

Appendix: Calculation concerning the tangent slope control

We shall denote the voltages between the lowest line in fig. 10 and the points F , P , Q , K and G by V_F , V_P , etc.

The calibration of the scale of R_5 is carried out with $V_Q = V_P = V_F = V_G$, so that at every position of the slider of R_5 the relation between V_K and the value I_{Cal} of the cell current as shown on the calibrated scale is

$$V_K + I_{Cal}R_3 = V_G \dots \dots \dots (1)$$

The value of I_{Cal} so obtained is thus only equal to the true value I of the cell current if $V_Q = V_G$. If $V_F \neq V_G$ and the slider of R_{13} is not positioned completely to the right, then this condition is not fulfilled.

In carrying out a measurement V_1 is first so adjusted that, whilst $V_Q = V_P = V_F = V_G$, the cell voltage V is equal to the half-wave potential V_s . From fig. 10 we read that in this case the following holds:

$$V_1 - V_F = V_s \dots \dots \dots (2)$$

Maintaining the relationship (2), a change is now made in V_F . We now call the cell current and cell voltage I' and V' respectively (see fig. 9), whilst the slope of the tangent at the point under consideration (A) on the polarogram is indicated by α .

The sliders of R_5 and R_{13} are next so adjusted that the following relationships hold:

$$V_P = V_Q \dots \dots \dots (3)$$

$$\frac{\partial V_P}{\partial V_F} = \frac{\partial V_Q}{\partial V_F} \dots \dots \dots (4)$$

We shall now show that an adjustment of R_5 is thereby reached at which V_K has such a value that the value of I_{Cal} obtained from (1) is equal to the current indicated in fig. 9 by I_3 :

If we consider first the branch $R_{12} - R_{13}$ we can write for V_Q :

$$V_Q = V_G + (V_F - V_G) \frac{R'_{13}}{R_{12} + R_{13}} \dots \dots (5)$$

thus,

$$\frac{\partial V_Q}{\partial V_F} = \frac{R'_{13}}{R_{12} + R_{13}} \dots \dots \dots (6)$$

For V_P , according to fig. 10, we can write.

$$V_P = V_K + I'R_3 = V_1 - V' \dots \dots \dots (7)$$

Since a variation of V_F causes an equally great variation in V_1 , $\partial V_P / \partial V_F$ is equal to $\partial V_1 / \partial V_F$. Considering now the branch $Z - R_3$ and bearing in mind that for voltage variations the cell Z represents a resistance $\cot \alpha$, then according to fig. 10 the following applies:

$$\frac{\partial V_P}{\partial V_F} = \frac{R_3}{R_3 + \cot \alpha} \dots \dots \dots (8)$$

To satisfy (3), V_K and R'_{13} must be so adjusted, according to (5) and (7), that the following equation is satisfied:

$$V_K = V_G + (V_F - V_G) \frac{R'_{13}}{R_{12} + R_{13}} - I'R_3 \dots \dots (9)$$

At the same time, to satisfy (4), R'_{13} according to (6) and (8) must be equal to the value derived from

$$\frac{R'_{13}}{R_{12} + R_{13}} = \frac{R_3}{R_3 + \cot \alpha} \dots \dots \dots (10)$$

(V_K plays no role in this).

We find now for the value of I_{Cal} corresponding to V_K , from (1), (9), and (10):

$$I_{Cal} = I' + (V_G - V_F) \frac{1}{R_3 + \cot \alpha} \dots \dots \dots (11)$$

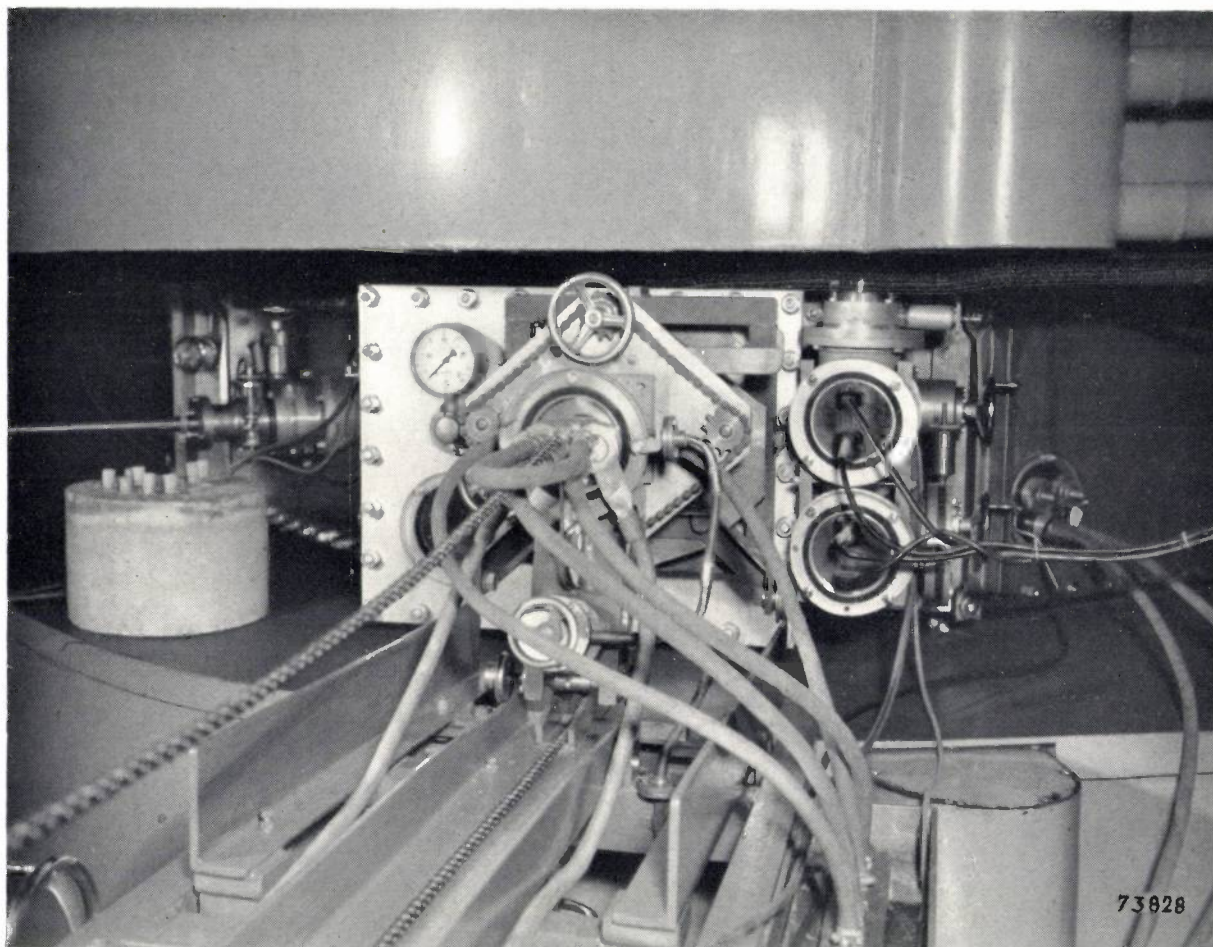
Since it follows from (3), (5), (7) and (10) that:

$$V_s - V' = (V_G - V_F) \frac{\cot \alpha}{R_3 + \cot \alpha},$$

we can write for (11):

$$I_{Cal} = I' + (V_s - V') \tan \alpha = I_3.$$

Summary. In this article a polarograph is described which is direct reading, so that in making a quantitative analysis it is not necessary to plot a complete polarogram. The concentration of certain metal ions in the solution to be analysed can be deduced from the difference of two readings on the instrument. As cathode, use is made of a mercury drop which is renewed every 1 to 4 seconds. The current flowing in the solution is measured at the moments just prior to the detachment of the drops. As an indicating instrument a cathode ray oscilloscope is used. The possibility of reading the value of the current steps directly on the instrument is obtained by use of a circuit with which the graphical construction on the polarogram is simulated electrically.



THE SYNCHROCYCLOTRON AT AMSTERDAM

IV. DETAILS OF CONSTRUCTION AND ANCILLARY EQUIPMENT

by F. A. HEYN *) and J. J. BURGERJON **).

621.384.61

The Philips synchrocyclotron at Amsterdam has now been in continual use for three years as a tool for nuclear-physics research and the production of radio-active isotopes. During these three years many other projects of a similar kind have been undertaken or designed in other countries and, although a number of these may have left the Amsterdam equipment far behind as regards the attainable particle energy (as well as in size and cost), equipment such as the Amsterdam synchrocyclotron continues to fulfil an important function. In its own class this synchrocyclotron can be regarded as one of the best at present in operation, its special features being the very high beam current (and hence a high yield of radio-active substances) and its high degree of reliability in service.

Introduction

Three articles have already appeared in this Review on the subject of the Philips synchrocyclotron at the Institute for Nuclear Physics Research at Amsterdam ¹⁾. In these articles descriptions were

given of the general scope and dimensions of this equipment, followed by details of the design of some of the essential components such as the oscillator, the modulator and the electromagnet. In the present article we propose to complete the series by saying something about the other main components and ancillary gear. Although numerous detail problems, were solved without fundamental knowledge — some, indeed, purely empirically — a description of them may well be useful, having regard to the fact that it is often just such details which

*) Now professor at the Technical University, Delft.

***) Now attached to the South African Council for Scientific and Industrial Research at Pretoria.

¹⁾ F. A. Heyn, The synchrocyclotron at Amsterdam:
I. General description of the installation,
II. The oscillator and the modulator,
III. The electromagnet,
Philips tech. Rev. **12**, 241-247, 247-256, and 349-364, 1951
(hereinafter referred to as I, II and III).

decide the success of a project. It may well be said, for example, that the high beam current in this synchrocyclotron (a good $20 \mu\text{A}$) has to a large extent been achieved as a result of the empirical design of the ion source. The high degree of reliability, and in particular the

huge electromagnet, deuterium ions from the ion source describe circular paths in the Dees in such a manner that they are accelerated by the alternating voltage each time they pass the gap between the Dees, their orbits becoming larger and larger with each revolution. When some 5000 revolutions have been completed the deuterons have acquired an energy of 28 McV and have then reached the region of

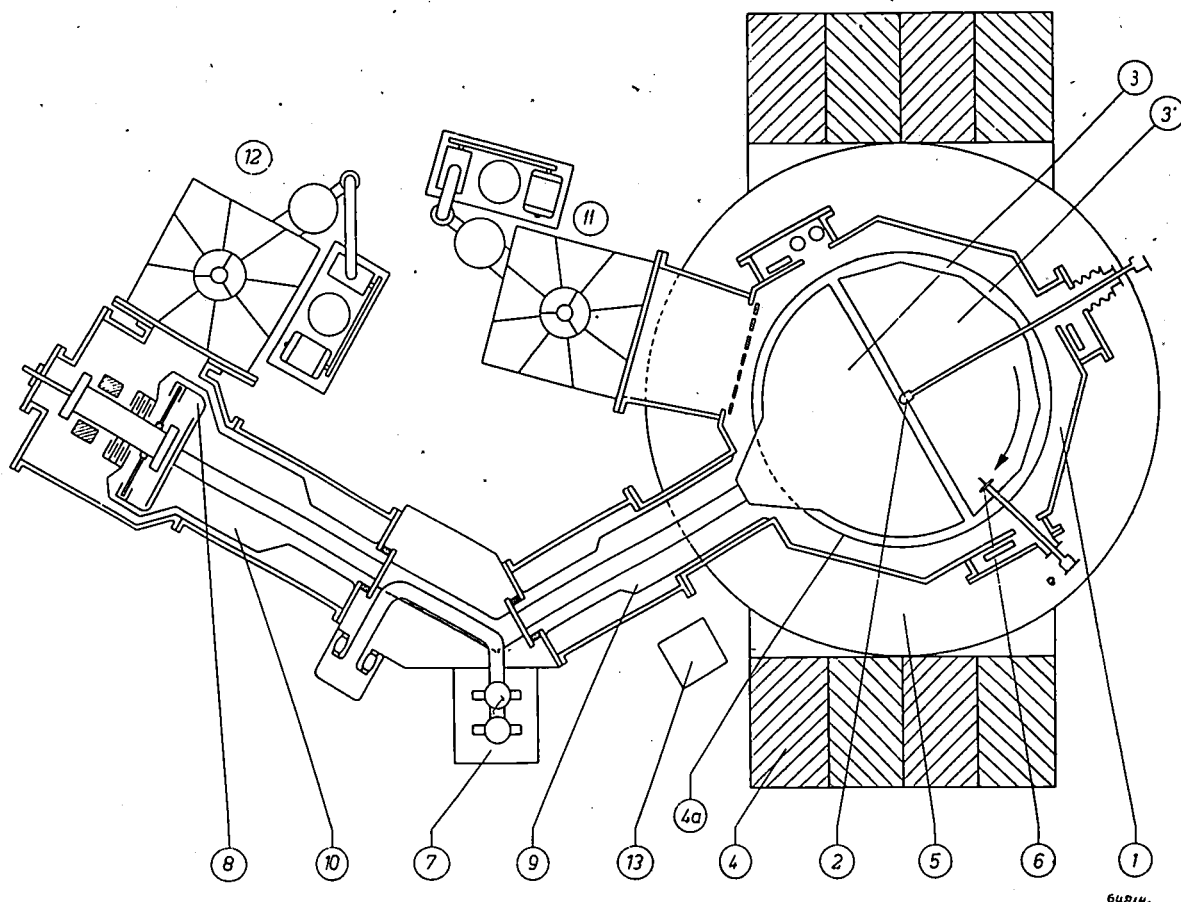


Fig. 1. Diagrammatic cross-section of the Philips synchrocyclotron at Amsterdam. 1 evacuated acceleration chamber; 2 ion source; 3 and 3' Dees; 4 magnet (4a face of magnet pole); 5 energizing coils; 6 target; 7 valves and other components of the oscillator; 8 modulator; 9 and 10 coaxial transmission line; 11 and 12 vacuum pumps; 13 boron ionization chamber for monitoring the radiation emitted.

small amount of time wasted owing to interruptions during operation, may be ascribed to the unconventional design of the acceleration chamber and its associated components (which allow of rapid dismantling), the efficient construction of the vacuum system and the numerous safety devices provided to avoid errors in manipulation.

For the convenience of the reader a brief survey of the general design of the cyclotron is given, with reference to *fig. 1*. Inside the circular, evacuated acceleration chamber the two Dees are mounted, between which a high frequency voltage of about 10.7 Mc/s with a peak of roughly 14 kV is applied²⁾. Under the influence of the magnetic field of the

the periphery of the acceleration chamber. Here they strike the target, upon which the material to be irradiated is placed. The oscillator generating the high frequency voltage, and the transmission line by means of which this is applied to the Dees are also visible in the diagram (*fig. 1*). The function of the modulator need not be discussed here; we refer the reader instead to I and II.

The acceleration chamber

The design of the synchrocyclotron was based on a maximum particle-orbit radius of about 80 cm, with a pole-piece radius of 90 cm. To suit these dimensions the acceleration chamber had to be roughly 2 metres in diameter. This chamber is evacuated to a very low pressure (about 10^{-4} mm Hg) to prevent loss of accelerated particles owing to collision with gas molecules.

²⁾ At the present time the cyclotron is operated at 20 kV, so that the particles describe a smaller number of revolutions.

The top and bottom of the chamber have thus to withstand a pressure of something like 30 tons, and they can be supported only at the sides, since

supports in the centre would obviously intercept the particles in their course. The best solution involves using the poles of the magnet themselves as top and bottom of the chamber. The large width of the magnet gap — which is responsible for the greater part of the 500 000 ampere turns required (see III) — is thus utilized to the greatest possible extent. This is important since the Dees should be given the greatest possible depth in order to minimize the loss of particles which oscillate, with considerable vertical amplitude in their initial revolutions and tend to collide with the top and bottom surfaces of the Dees.

To facilitate the "shimming" of the magnetic field (see III) and, more especially, to facilitate assembly of the acceleration chamber, in the final design 2-centimetre air-gaps were left in each pole, 7 cm from the face. Thus, the chamber in fact has separate top and bottom plates, each of these being 7 cm thick; these, however, are of the same material as the magnet itself. Being so thick, they hardly deflect at all under the atmospheric pressure. The supports at the sides are arranged symmetrically so that any buckling will also be symmetrical. The radial variation of the magnetic field, whose adjustment is very critical (see III), has been found to be undisturbed by the very slight buckling which occurs.

In order that the magnetic field shall not be affected in any way, it is essential also that everything between the top and bottom faces be made of non-magnetic material, including the vertical walls. Usually, the walls are made in one piece, of brass or non-magnetic steel and on which the top and bottom plates are laid, the assembly then being wheeled between the poles, or welded together when in position. Here we have tried out a new form of construction. The chamber is built up from elements of convenient size, so as to be easily demountable. The main components are: two cast brass rings with octagonal circumference, laid close around the top and bottom plates; eight brass pillars placed between the plates round the periphery and to which the rings are screwed; and eight rolled brass plates fitted to the sides of the rings. The chamber was built up from these elements directly between the poles. Airtight joints were obtained by means of rubber gaskets (see *figs. 2 and 3*). It should be pointed out here that, for the chamber to be demountable, crossings of rubber seals are unavoidable. Surprisingly enough, this works quite well, provided that each crossing involves one round and one flat rubber (not round-and-round or flat-and-flat). The total leakage into the acceleration chamber

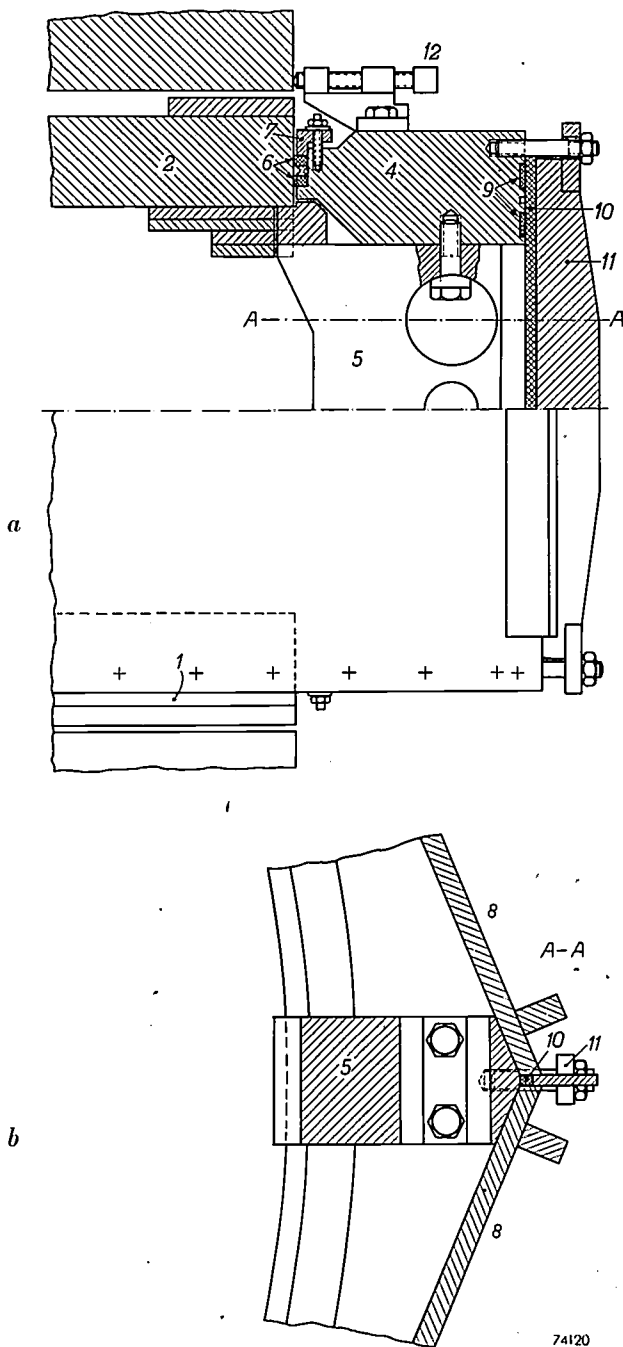


Fig. 2. One of the eight corners of the acceleration chamber: a) vertical cross-section, b) horizontal cross-section. 1 bottom plate; 2 top plate of the chamber; 4 upper brass ring; 5 brass pillar. The space between the rings and plates is sealed, vacuum-tight, by round-section rubber rings 6 pressed into a groove by packing clamps 7. Two rubber strips 9 ensure a vacuum-tight joint between the brass rings and side plates 8. The spaces between the side plate, which are 8 mm, are closed by means of 10 mm rubber cord 10, held in place against the pillar by a heavy clamp 11. This cord crosses the horizontal strips 9, but proper matching of strip and cord, together with sufficient pressure, ensures an air-tight joint. 12 set-screw for centering the acceleration chamber between the magnet poles.

built up in this way, as well as that of the connected spaces (oscillatory system for the Dee-voltage, pump lines etc.) is less than 10^{-5} litre (at normal atmospheric pressure) per second, that is, about 6 litres per week for a total volume of 2 cubic metres, which is quite acceptable.

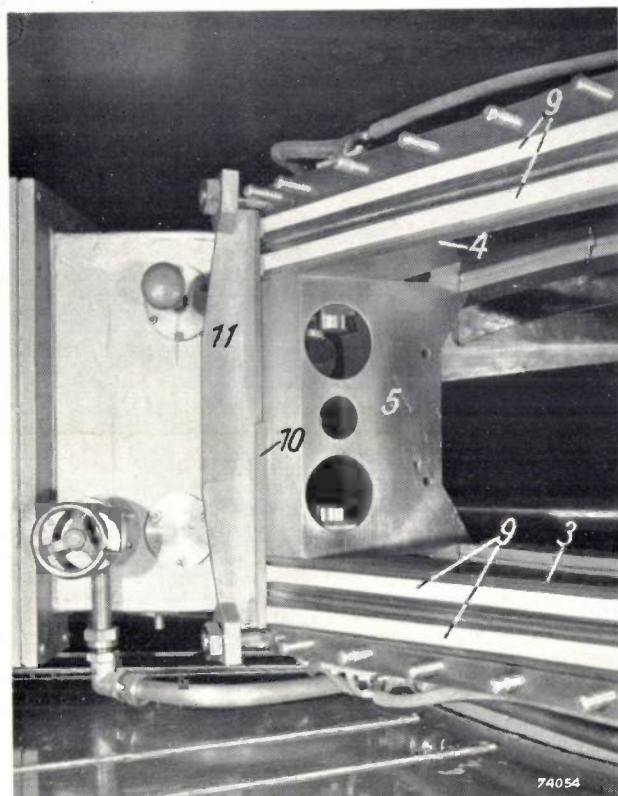


Fig. 3. A corner of the acceleration chamber with side plate removed, showing brass pillar 5 (see fig. 2) between the upper and lower brass rings (3 and 4), and the horizontal rubber strips (9), the rubber cord (10) (top half cut away to show detail), and the associated packing clamp (11).

The acceleration chamber is supported on the bottom pole by six brass spacers, 2 cm in thickness. When the magnetic field is applied it tends, by reason of the barrel-shaped pattern of the lines of force in the magnet gap (see III), to pull up the top plate against the forces of gravity and air-pressure. Since the top and bottom plates are able to move slightly on their seals, this displacement does not result in leakage; however, to limit the amount of movement, six brass blocks are included between the top plate and the upper pole, the thickness of these blocks being calculated to suit the pole spacing with the magnetic field in operation. In fact, owing to the mutual attraction of the poles, which causes the magnet-yoke to give slightly, the spacing in question is somewhat smaller than in the absence of the field. (This apparently insignificant constructional problem, which is solved quite simply by the

method of sealing employed, gave rise to much trouble in other cyclotrons.)

The inner faces of the top and bottom plates, as well as the shims at the edges (see III, fig. 4c) are clad with copper in order to reduce losses in the H.F. field of the Dees. Nevertheless, the heat developed as a result of losses is such that water-cooling must be employed.

The Dees

Of the two Dees in the acceleration chamber the one is earthed whilst the other carries the 14 kV H.F. voltage with respect to earth. The latter is suspended on a stem (9, fig. 1) which functions also as a transmission line; the end of this stem is clamped between two rings with "Kersima" insulators (see II) and each of these rings is held in place by four pins (fig. 4), by means of which the Dee can be adjusted so as to lie parallel to the earthed Dee. The gap between the Dees is 5 cm.

In view of the limited (though fairly high) compression resistance of the insulators and to avoid too much sagging, the Dee and its stem are fabricated from a light framework of brass, jacketed with copper. The top and bottom halves of the jacket are water-cooled; the cooling system is illustrated in fig. 4.

The height of the mouth of the Dee was made as large as possible, viz. 20 cm, in order to hold as many particles from the ion source as possible and, as already mentioned, to minimize the loss of particles owing to collisions with the top and bottom surfaces. It was not possible to use a larger mouth height in the Dees in conjunction with the given magnet gap, in view of the risk of flashover to the top and bottom plates of the chamber. Furthermore, the capacitance of the Dee should not be too high (see II). For the same reason the Dee decreases in height towards the periphery of the acceleration chamber, where the oscillation amplitude of the particles is smaller; at the point where the radius of the path is the greatest, the height of the Dee is 10 cm. The earthed Dee, which consists of plates attached to the top and bottom of the chamber, presents to the particles roughly the same amount of space as the other Dee.

The housing containing the stem of the Dee is mounted on a carriage and can be moved on rails, enabling this Dee to be easily removed from and re-introduced into the acceleration chamber; for this purpose two of the columns and two side plates must be removed.

The ion source

It has already been pointed out that a copious source of ions is necessary to ensure high beam intensity. This is all the more important since only a fraction of the ions produced are accelerated; acceleration of the particles can commence only at the beginning of each modulation cycle (see II) and, further, only those particles are accelerated which, in a certain phase of the Dee voltage, occur at a certain point, such that the phase stabilization operates in their favour.

Quite a number of designs were tried out for the ion source and the following was ultimately found to give very satisfactory results. The gas to be

with respect to the earthed walls, this voltage being applied through a current limiting resistor. Owing to the relatively high pressure in the chamber, a gas discharge occurs, as a result of which the tungsten becomes incandescent and emits electrons (an operating voltage of 210 V gives rise to a total current of 150 mA). On account of the presence of the intense magnetic field the electrons are able to move only in narrow helices around the magnetic

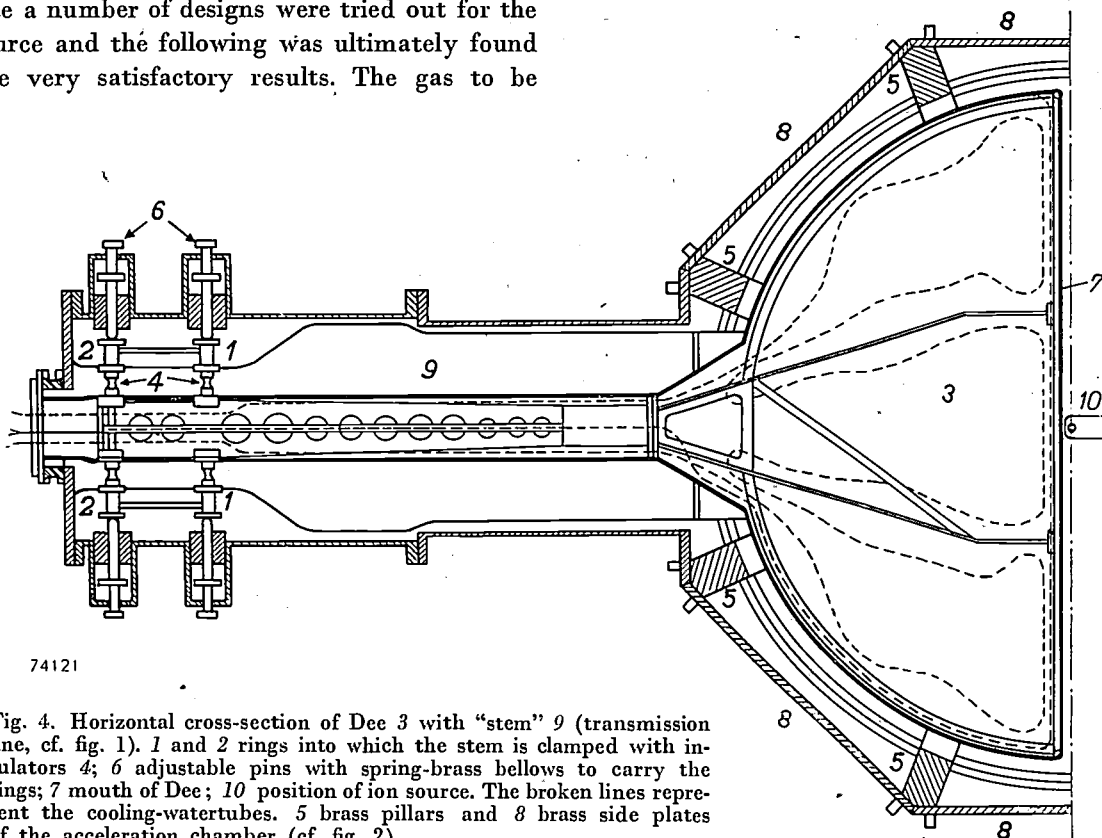


Fig. 4. Horizontal cross-section of Dee 3 with "stem" 9 (transmission line, cf. fig. 1). 1 and 2 rings into which the stem is clamped with insulators 4; 6 adjustable pins with spring-brass bellows to carry the rings; 7 mouth of Dee; 10 position of ion source. The broken lines represent the cooling-watertubes. 5 brass pillars and 8 brass side plates of the acceleration chamber (cf. fig. 2).

ionized, usually heavy hydrogen, is supplied to an earthed copper chamber located just above the gap between the Dees. This chamber is fitted with a nozzle, directed downwards in the direction of the magnetic lines of force (fig. 5) and through which gas will enter the accelerating chamber. The orifice of the nozzle is sufficiently long (60 mm) and narrow (5 mm diam.) to ensure considerable difference in pressure between the two extremities; with a flow of gas of 0.085 cc/sec (about 0.3 litre per hour referred to 76 cm Hg), the pressure inside the chamber of the ion source is 0.1 mm Hg, whereas the pressure just downstream of the nozzle is a good 10^{-4} mm Hg. Inside the chamber there is a horizontal tungsten pin, 4 mm thick, mounted in insulating material and carrying a voltage of -450 V

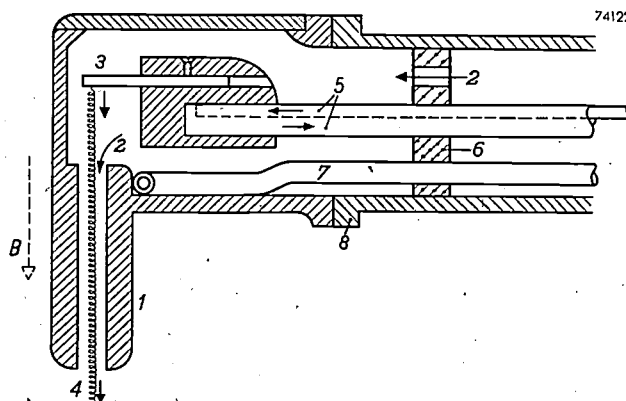


Fig. 5. The ion source. 1 narrow-bore nozzle, parallel to magnetic field B ; 2 gas inlet; 3 tungsten pin; 4 electron stream; 5 cooling tubes (also current conductor for the tungsten pin); 6 insulator; 7 cooling tube for walls of chamber; 8 flange to which the ion chamber is screwed.

lines of force which are parallel to the nozzle. Hence a highly concentrated electron beam issues from the nozzle and passes through the acceleration chamber, arriving finally on a water-cooled plate

cathode, and so on. The electrons, thus passing repeatedly to and fro through the nozzle, produce sufficient ions there to maintain the tungsten cathode at a high temperature by the ion bombardment,

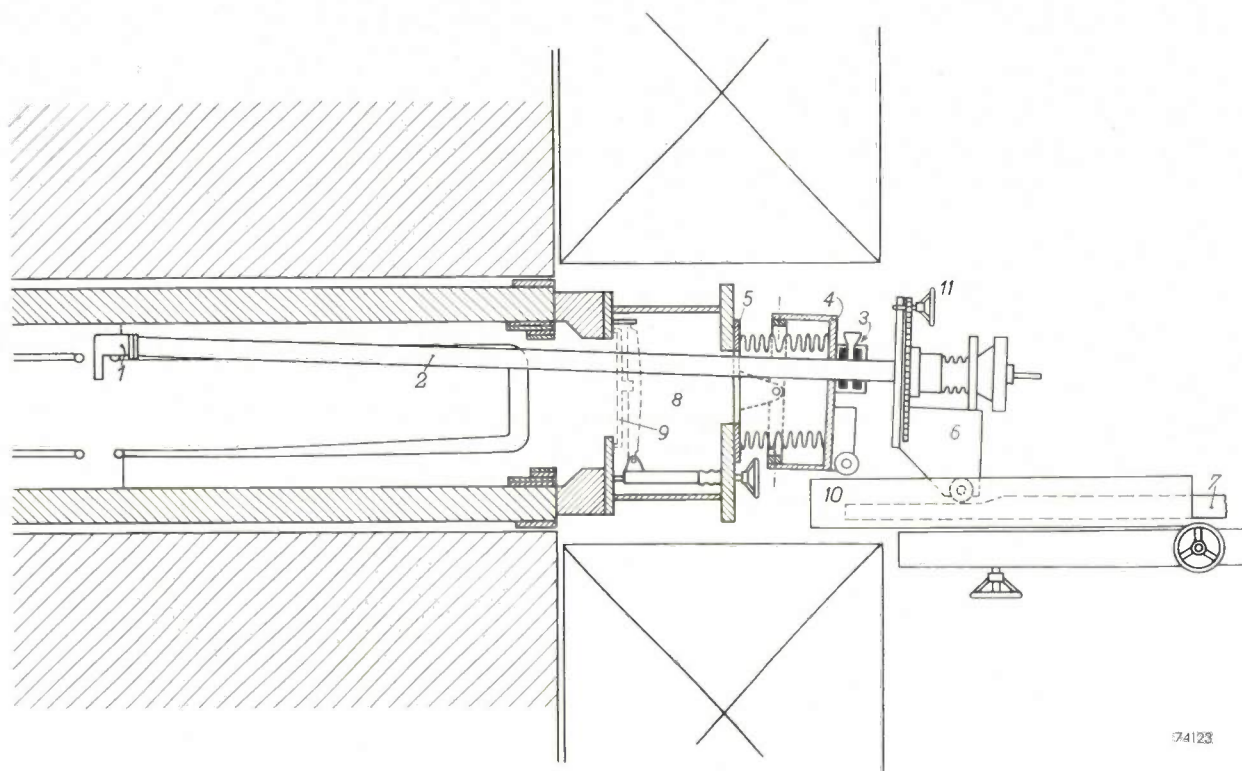


Fig. 6. Holder and controls of the ion source. 1 ion chamber with nozzle; 2 tube carrying ion chamber, containing insulated tube for the cathode and cooling tube for the chamber walls. The end of the long tube is supported by a bearing on carriage 6. Plate 4 is another bearing in which the tube slides through a vacuum-tight gland 3 (consisting of two rubber seals with vacuum oil between them). The ion source can be inserted more or less deeply into the acceleration chamber by means of a handwheel 11. Further, plate 4 can be tilted in two directions with respect to the flange 5 to which it is attached by spring-brass vacuum-tight bellows, enabling the ion source to be moved up and down or sideways in the Dee gap. That position of the ion source which yields the highest beam current can thus be found while the cyclotron is operating. In order to withdraw the ion source from the acceleration chamber, for example to renew the tungsten pin, the carriage 6 is pulled back along the rails 7. When the ion source reaches the vacuum lock 8, the cover 9 is closed (see fig. 8) and the flange 5 can then be loosened; the flange, with plate 4 and the whole ion source assembly, is then retracted on the rails 10 which are first raised so that the wheels on plate 4 come to rest on them.

at the bottom of the chamber. The electrons spiralling downwards have to cover a considerable distance, and they thus produce in the acceleration chamber a very large quantity of ions, even though the gas pressure there is low.

It is not yet possible to explain in every detail the action of this ion source (which might better be referred to as an electron source). It is worthy of note that the gas discharge in the source is extinguished immediately the H.F. Dee voltage is cut off; furthermore, in the absence of this H.F. voltage, it is not possible to start up the source at the above-mentioned ignition potential of 450 V. Probably, the explanation is that, in the negative phase of the Dee voltage, the electrons are reflected from the Dee field, thereby resume a helical path upwards through the nozzle, are again reflected by the

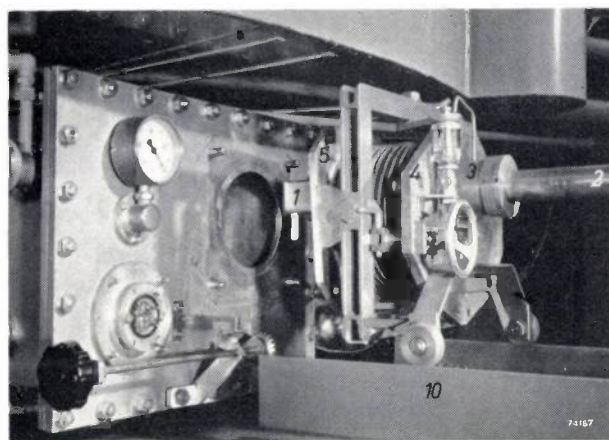


Fig. 7. The ion source fully retracted. In the photograph will be seen: 1 the ion source chamber; 2 tube, 3 gland; 4 plate and 5 flange on the bellows; 10 rails on which the wheels on plate 4 run.

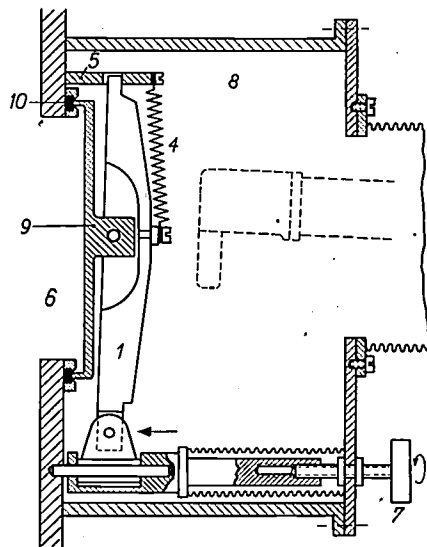
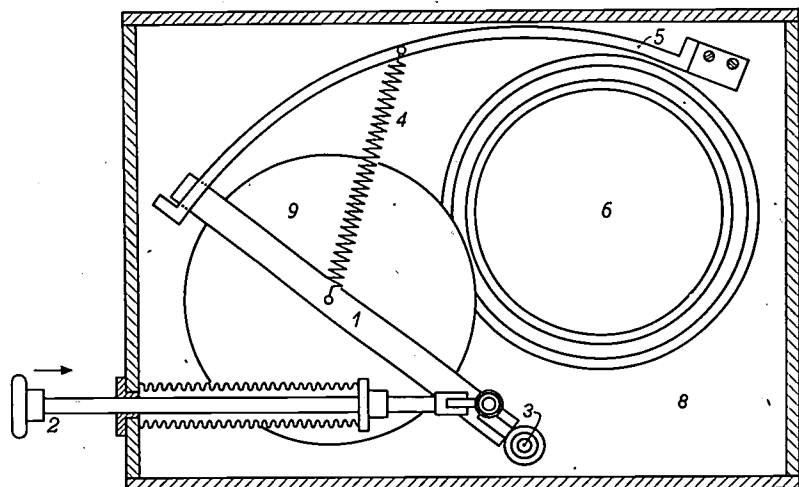


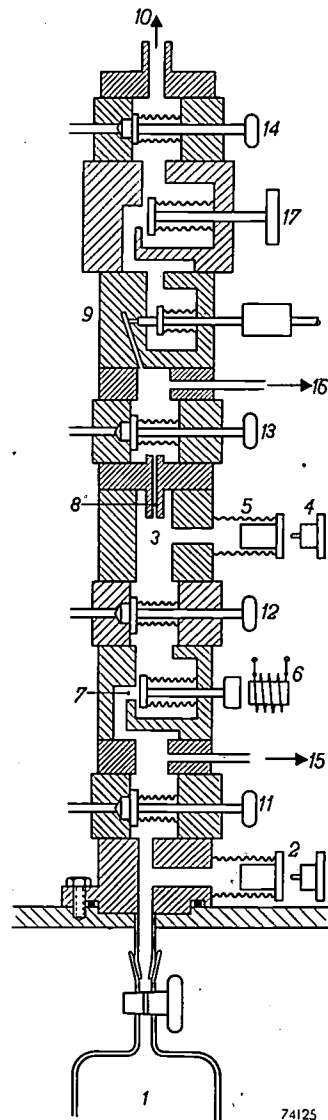
Fig. 8. The vacuum lock for the ion source. Cover 9 (cf. fig. 6) is mounted on an arm 1 which, when the pin 2 (with bellows) is pushed in, swings on a pivot 3. This demands little effort as the movement is assisted by a spring 4. When the arm is rotated, the extremity slides on a track 5 until it reaches a stop, in which position the cover is exactly in front of aperture 6 in the acceleration chamber (through which the ion source has already been drawn; this is indicated by the broken lines on the right-hand side of the cross-section). When the handwheel 7 is turned, the lower end of the arm 1 is pushed inwards by a screw spring-bellows, so that the cover 9 is pressed firmly against a packing ring 10. Air can then be admitted to the trap.

i.e. an independent discharge occurs. In order to ensure the emission of sufficient electrons to sustain this process (and to provide the necessary heavy ionization within the acceleration chamber) in spite of the large space-charge in the nozzle, the nozzle must not be too narrow. Experience has shown, in fact, that the discharge cannot be established with, for example, a 4 mm nozzle.

The tungsten pin is held in a water-cooled copper block (fig. 5). The cooling tube serves simultaneously as support and as conductor for the feed to the pin. To prevent the extremity of the pin from being cooled too much, this pin must be of a certain minimum length. The walls of the ion source are likewise water-cooled.

Outside the acceleration chamber the tube carrying the whole ion source assembly is carried on a plate which is adjustable in two directions, in such a way that the most satisfactory position for the source can be found when the cyclotron is in operation. The tungsten pin has to be renewed every other day, since the yield of ions drops owing to melting and evaporation of the tungsten; to facilitate replacement, a vacuum lock is incorporated through which the source can be withdrawn without admitting air to the acceleration chamber. The mechanical design may be seen from figs 6, 7 and 8, in whose subscripts further details are given.

Fig. 9. Valve assembly for continuous feeding of the ion source with gas (heavy hydrogen or helium) at constant pressure. 1 storage bottle; 2 mechanism for giving signal in control room as soon as pressure in the storage bottle drops below 10 cm Hg, i.e. warning that bottle must be renewed; 3 buffer volume in which pressure is kept practically constant; when pressure drops, the switch 4, operating through spring-brass bellows, actuates an electromagnet 6 which opens the inlet valve 7 a little more; 8 duct offering high resistance to flow of gas, to damp out fluctuations; 9 needle valve to regulate gas supply to the ion source (operated from control room by means of a servomotor); 10 gas line to ion source; cocks 11, 12, 13 and 14 are used for withdrawing air that enters when the gas storage bottle, or the tungsten pin in the ion source is renewed; 15 and 16 are connections for pressure gauges; 17 stop cock.



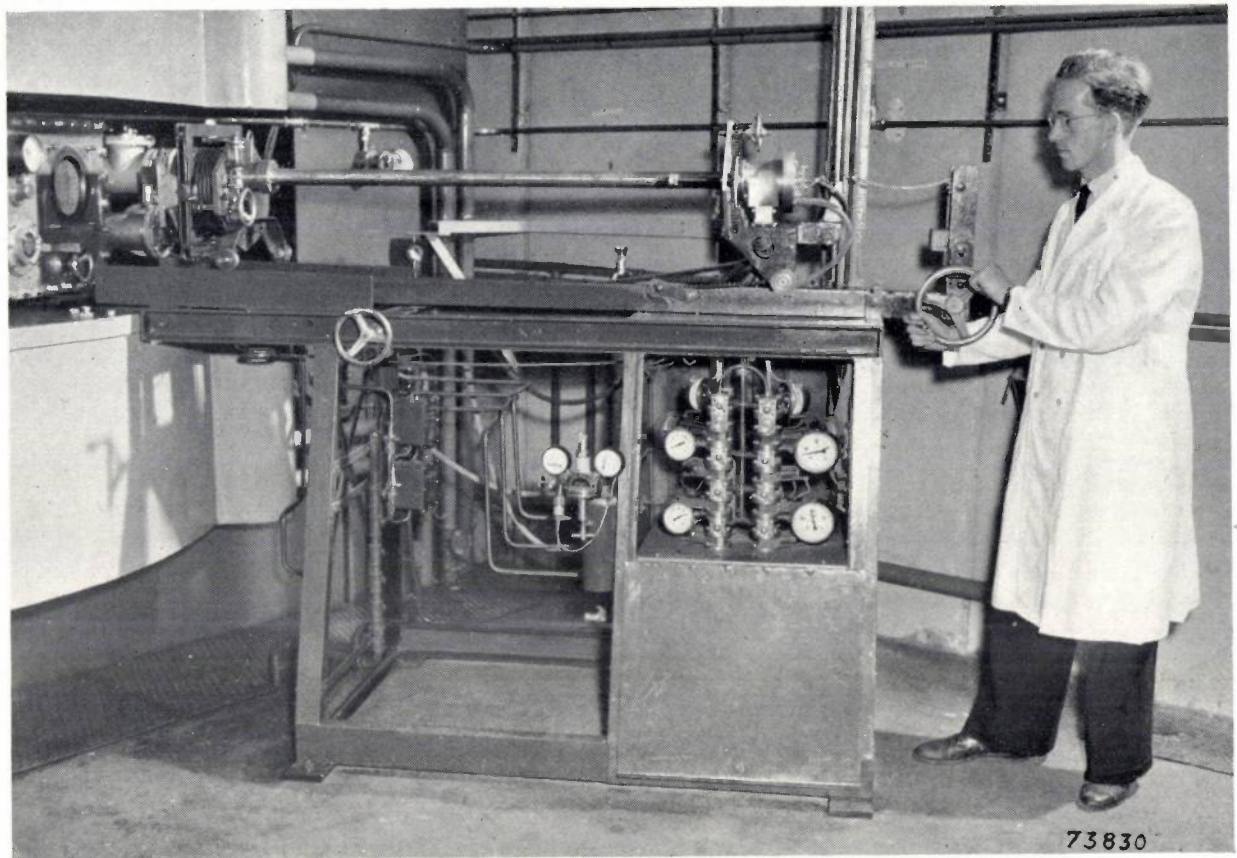


Fig. 10. The complete apparatus associated with the ion source. The source is shown withdrawn from the chamber. Below, in the framework supporting the rails, is the gas supply apparatus, which is duplicated for the acceleration of helium or deuterium ions as described. The two glass storage bottles are housed in the compartment below. The consumption of gas is 0.3 litre per hour (referred to 76 cm Hg).

The heavy hydrogen, which has to be continuously supplied to the ion source, is obtained from heavy water by electrolysis and is stored in glass bottles, the pressure in the chamber of the ion source being regulated by means of a needle valve. To avoid the necessity of constant readjustment of the valve while the bottle is emptying, a buffer volume is included between the valve and the bottle, in which the pressure is automatically kept constant. The buffer volume contains a spring-brass bellows which contracts when the pressure drops, and trips a switch; this, with the aid of an electromagnetic control, opens the valve to the buffer volume more; the pressure then rises to the normal level and the expanding bellows once more reduces the gas supply. This "discontinuous" control of the pressure does certainly produce in the buffer volume small variations about a pre-set, average, value, but owing to the high resistance encountered in a narrow duct between the buffer volume and the valve, these variations are not passed on to the ion source to any noticeable extent.

Fig. 9 is a cross-section of the gas supply system,

showing sundry other components; *fig. 10* is a photograph of the complete equipment associated with the ion source.

The target

The beam of accelerated particles, spiralling outwards, comes more and more into focus and finally strikes the target, on which the substance

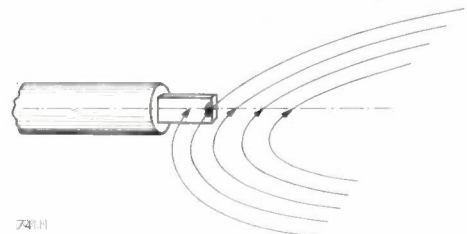


Fig. 11. Location of the target with respect to the beam of accelerated particles. These strike the target mainly within the area shown hatched.

to be bombarded has been placed; the manner in which this takes place is illustrated diagrammatically in *fig. 11*.

At the target the great heat generated constitutes

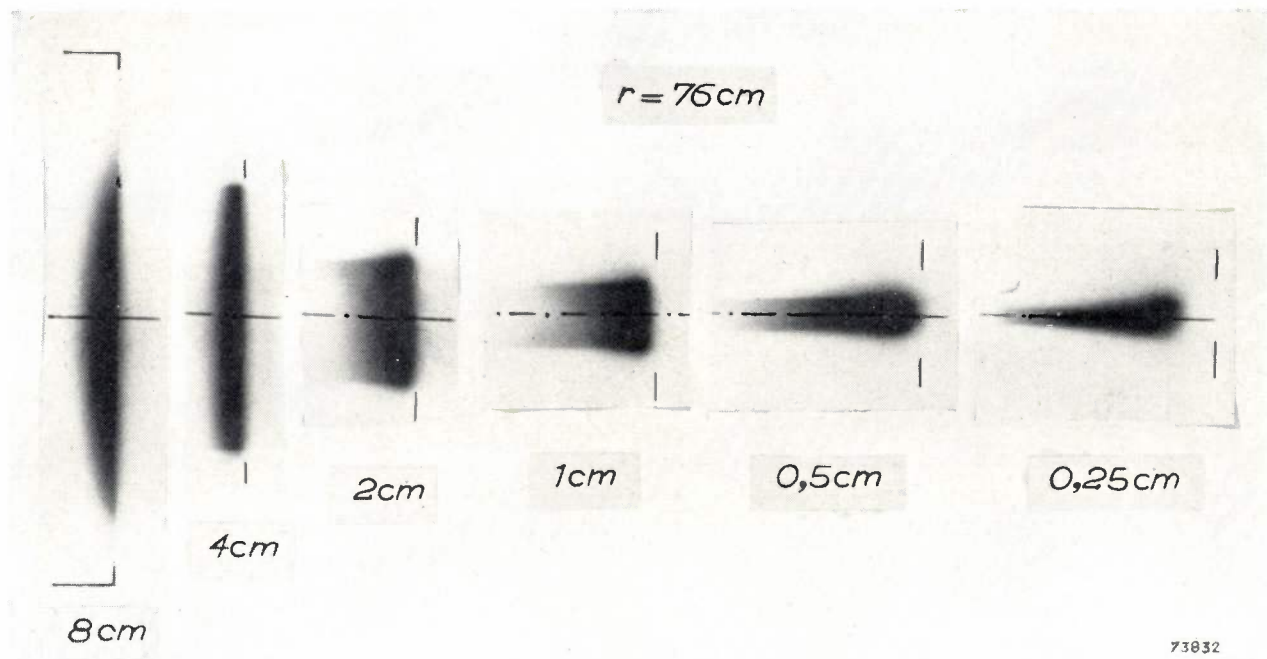


Fig. 12. Distribution of intensity of the beam on the target, with targets of different vertical dimension, at a radius r of 76 cm in the acceleration chamber. The lines indicate the position of the edge of the targets. These illustrations are contact prints on a photographic film, of a set of copper plates, each of which was placed in the beam for a few moments (autoradiograms). With the plates of small height, the prints show considerable over-exposure; in the case of the 0.25 cm plate, which was not easy to cool, the front end has melted off.

For more quantitative details of the beam distribution, the activity is measured at a number of points on the activated plates with a Geiger counter.

quite a problem; with a particle energy of 28 MeV and a beam current of $20 \mu\text{A}$, the beam dissipates 560 W at the target, a large part of the power being concentrated within the hatched area shown in fig. 11. This repeatedly gave rise to evaporation and melting of the substance irradiated and efficient dispersal of the heat became an all-important problem.

If the beam were confined to the median plane of the acceleration chamber (and if the component sections of the path of the particles were truly circular), all the particles would be intercepted by the edge of the target where this is intersected by the median plane. Actually, however, the vertical spread of the beam is considerable, in consequence of the vertical oscillation of the particles. Many particles therefore shoot past the target, above or below it, perform one or more extra revolutions, thus acquiring a larger path radius, and finally strike the target at points away from the tip. (A similar effect is produced because, owing to radial oscillation, each particle does not describe an expanding circle, but an expanding and precessing ellipse). If a longer target (in the vertical direction) is used, most of the particles will be intercepted at the edge, giving a high, narrow impact area; if the target is made any smaller in the vertical direction, more particles will be able to continue on their spiral path and the impact area will be low and wide, as will be seen from fig. 12. For the production of radio-active substances a uniform field is most favourable and a target of average height is therefore used.

It is also interesting to note the modification in the form

of the beam when the radius is varied. This modification is illustrated in fig. 13, from which it will be seen that the beam is not invariably symmetrically disposed with respect to the median plane; also that the height decreases when the radius is increased. (With a sufficiently small radius, i.e. much lower particle energy, the beam covers the whole height of the Dees³), the beam current as measured there being some hundreds of microamps.) Particles can be detected beyond even the normal working radius of 78 cm.

The heat developed at the target is carried away by a water-cooling system. Often the substances to be bombarded are made into high melting-point compounds (glasses) which are applied in the form of a thin layer to a copper water pipe. Where this is not possible, a block of metal is soldered to a water pipe and holes of, say, 1 mm diam. are drilled in the block, into which the substance in question is introduced. Metals to be bombarded are clamped between two water-cooled plates. A number of targets of different kinds is depicted in fig. 14; the long tubes for the water-cooling serve at the same time as holders when the radio-active or "hot" substances are removed or further handled. The target is changed in much the same way as the ion

³) It has been noticed that the Dees themselves become radioactive at certain points, which indicates that they are struck in these points by parts of the beam.

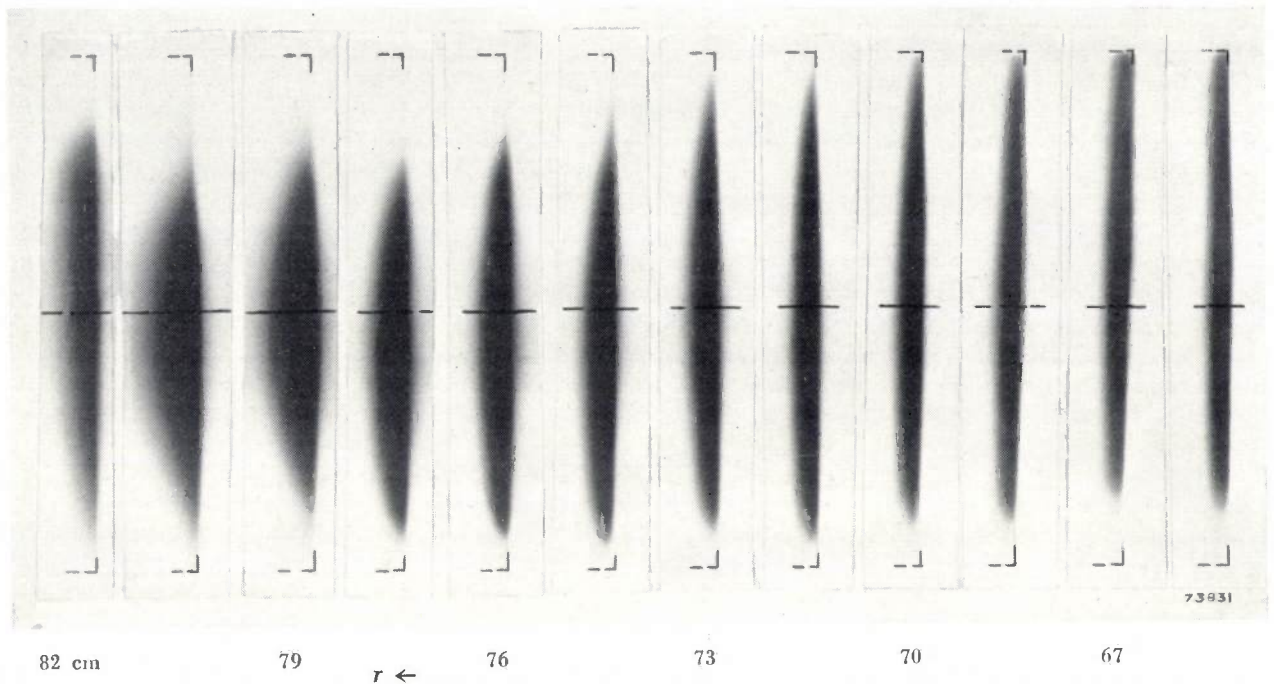


Fig. 13. Exposures of the beam as in fig. 12, obtained with one target 8 cm in height placed at different orbit radii r .

source, with the aid of a vacuum lock. This is very similar in design, except that all the controls are now at the rear, since personnel always have to keep at a safe distance from the "hot" material (fig. 15).

The depth of penetration of the target into the acceleration chamber can be increased or decreased so that it is struck by particles of lower or higher

energy, and the actual position is indicated by a scale, calibrated in MeV. The energy can also be checked by measuring the range of the particles: Several thin targets are then placed one behind the other; the higher the particle energy and therefore the greater the number of particles passing through the first targets in the series, the greater

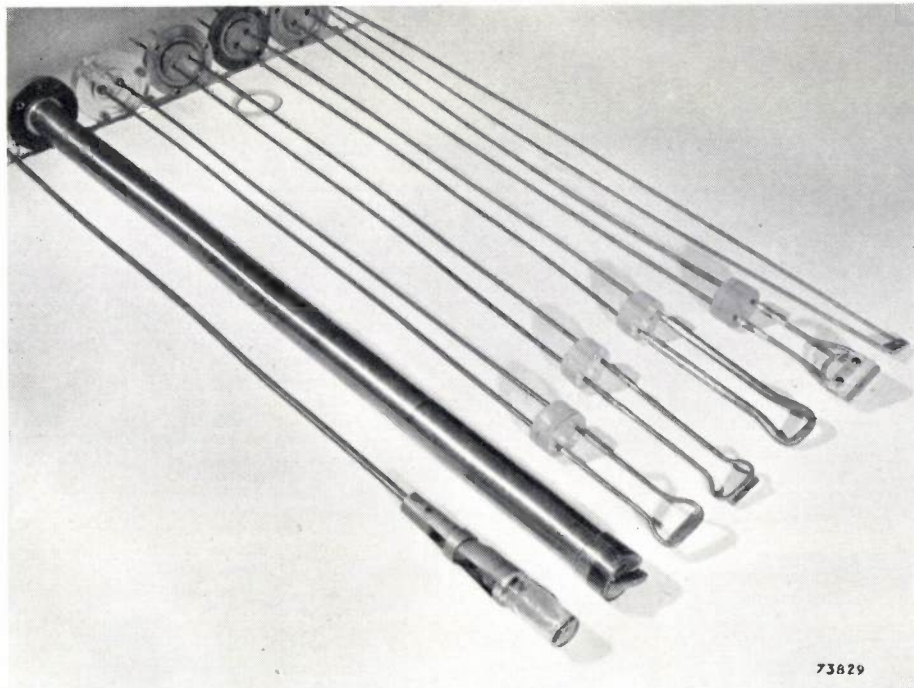


Fig. 14. Different types of target. These are handled by the long tubes which serve for the water-cooling.

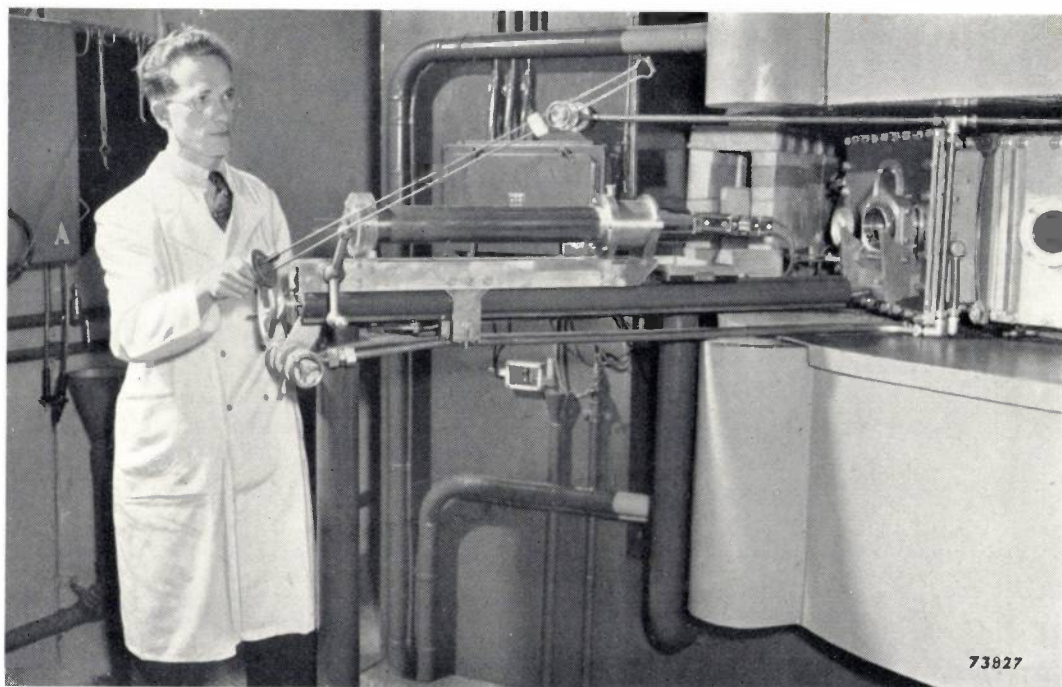


Fig. 15. The target is changed through a vacuum lock similar to that provided for the ion source; all manipulations are such that the operator remains at a safe distance from the "hot" target.

will be the effect produced at the later targets.

For regular production of radio-active substances it is important that the beam current be monitored continuously. This is effected in the equipment under discussion by measuring the temperature rise of the cooling-water at the target (fig. 16). The flow of water is kept constant by letting it run from a constant head tank. A multiple thermojunction is incorporated in the outgoing and return pipes and the thermo-current, which is a direct measure of the heat generated at the target, may be registered continuously by an instrument in the control room of the installation. In this way a direct record may be obtained of the performance of the cyclotron (the

number of μA -hours) during a given period. Since the flow-resistance of the whole cooling system is likely to vary, e.g. owing to differences in the form of the target, a calibration of the current measurement is necessary for each target. To carry out this measurement, a spiral of constantan wire is mounted in the cooling-water pipe, and a stabilized voltage can be applied so as to dissipate a known amount of power in the water.

In order to determine the performance of the cyclotron itself accurately, the effect of thermal radiation at the target must be eliminated. For this purpose, a thick copper target is used whose temperature hardly rises at all since its thermal conductivity is high.

The vacuum pumps

We have seen above that, to ensure a high beam current, a relatively high gas pressure is required in the ion source, together with a low pressure (about 10^{-4} mm Hg) in the acceleration chamber; in addition, the duct between the ion source and the acceleration chamber must not be too narrow. To satisfy these conflicting requirements the acceleration chamber has to be connected to a very high speed vacuum pump. At the same time the pump has to cope with the air-leaks which were mentioned above (and which could not be estimated accurately in advance) and it has to pump away any gases that may be liberated from the large metallic sur-

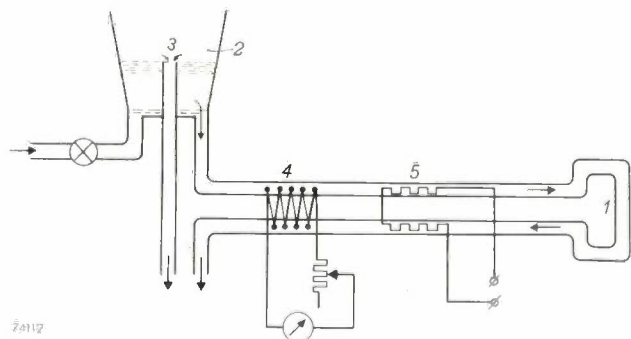


Fig. 16. The target 1 is cooled with water flowing from a reservoir 2 in which a constant head is maintained (3 is an overflow). The rise in the temperature of the cooling water is measured continuously by the multiple thermojunction 4, this rise being a measure of the beam current in the cyclotron. Calibration is possible by means of a constantan coil 5 which dissipates a known amount of power.

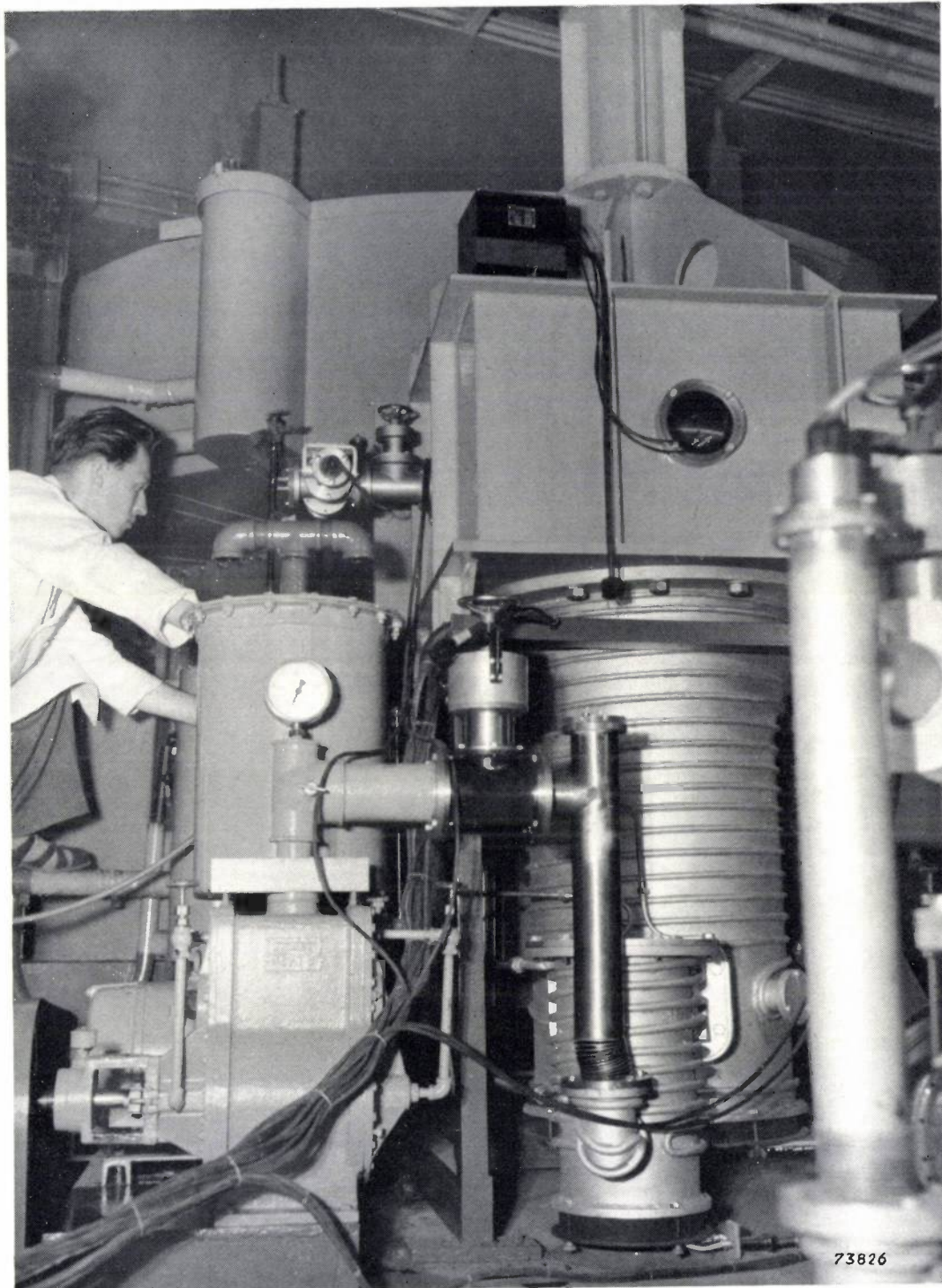


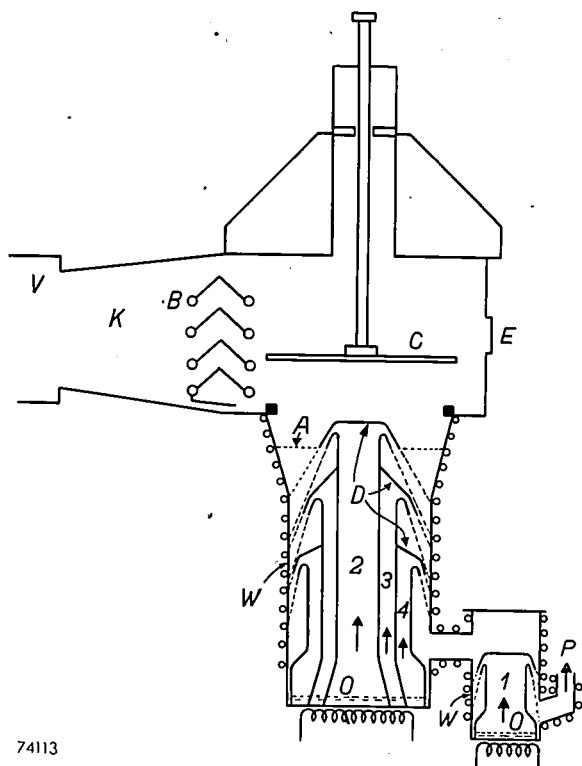
Fig. 17. One of the high-vacuum pumps of the cyclotron. On the right, below, will be seen the large cylinder comprising the 3-stage vapour diffusion pump and, in front of it, the booster pump, both with cooling-water tubes round the outside. On the left is the rotary backing pump. The valve and vapour trap are seen above the large cylinder and, behind these, the upper energizing coil of the magnet.

faces present. The general indication for the design of the pump was, therefore, that the pumping speed must be "as high as possible". Owing to unavoidable flow resistance, however, the pressure in the acceleration chamber, beyond a certain pumping speed is further reduced only very slightly

(see below). Carefully prepared estimates indicated that a pumping speed S_0 of 3000 litres per sec as measured at the pump should be aimed at. For comparative purposes it may be noted that normal good quality commercial mercury-vapour diffusion pumps have a pumping speed of say $S_0 = 50$ l/sec,

the speed of the largest vacuum pumps at present obtainable being 15 000 l/sec.

The pumping system designed by us, of which a photograph appears in *fig. 17*, comprises the usual backing pump and a 4-stage diffusion pump, viz. a booster and a 3-stage main pump. *Fig. 18* depicts the diffusion and booster stages diagrammatically.



74113

Fig. 18. Diagrammatic cross-section of the oil-vapour vacuum pump. *O* boiling oil; 1, 2, 3 and 4 chimneys of the four stages; *D* deflectors; *W* water-cooled casing. *P* backing line; *A* inlet aperture of the pump; *K* pump duct; *B* vapour trap; *C* valve. The plates of the vapour trap are cooled by a "Freon" refrigerator; they are tinned to promote reflection of the heat rays falling on them, and the whole assembly of plates is mounted on two thermal insulators. *E* inspection window, also to be seen in *fig. 17*. *V* acceleration chamber of the cyclotron.

At the bottom of the pumps there is a quantity of oil (Silicone Fluid DC 703) which is raised to boiling point by a heater. The vapour from the boiling oil rises through four "chimneys" in which the pressure is some tenths of a millimetre of mercury and deflectors direct the vapour through small annular outlet slots. From the chimney of each stage, therefore, a stream of vapour passes obliquely downwards. The walls against which the vapour jets strike are well cooled with water and the vapour is thus practically all condensed; accordingly, nearly all the vapour molecules move downwards at a high velocity (average value v_d). Gas molecules entering the pump by way of the inlet (*A*) and coming into contact with the vapour jets are drawn with the vapour; below the jets the gas is

thus "compressed", the jets forming a barrier between a zone of low pressure above and one of higher pressure below.

In *fig. 18* three other elements will be seen between the acceleration chamber and the pump proper, viz. a pump duct *K*, a vapour trap *B* and a valve *C*. The vapour trap serves in the first place to prevent oil vapour and the products of decomposition of the oil from entering the acceleration chamber; it consists of a number of metal plates in the form of inverted Vees, thermally insulated from the body and so arranged that no gas or vapour molecules can pass them without striking them at least once. The expansion tube of a "Freon" refrigerator is soldered to the plates and the temperature of these plates at the coldest points is -30°C ; any oil vapour is thus almost wholly condensed on the plates (the oil vapour pressure at the temperature in question is $< 10^{-6}$ mm Hg). Furthermore, the trap prevents an excess of water and other condensable vapour (grease and other impurities) from passing from the acceleration chamber to the pump and at the same time it assists in removing such vapours from the chamber. (Residual water and other vapours are gradually eliminated from the acceleration chamber at the commencement of each working period by means of a number of liquid air traps, connected to it.) The valve, which rests on a rubber seating, is used to close down the pump for the night, in order to prevent vapours from the pump or air from the low-vacuum volume from finding their way into the acceleration chamber⁴); this valve must not be opened until the pump is in full operation. On the other hand, closing of the valve also protects the pump in the event of an accident (breakage in any of the equipment connected to the chamber) that might allow large quantities of air to enter, for this has a deleterious effect on the fluid. The presence of the third component mentioned above, the pump duct (80 cm \times 50 cm cross-section), is purely a disadvantage, but is unavoidable, since the pump cannot be placed any closer to the acceleration chamber in view of the size of the energizing coils of the magnet.

The three components mentioned, together set up a certain flow resistance (referred to in our dis-

⁴) Small quantities of oil vaporize from the vapour trap during the night and enter the acceleration chamber, but these do no harm. This does, however, illustrate the reason why an oil vapour pump and not a mercury vapour pump should be used; mercury vapour penetrating into the acceleration chamber would do incalculable damage by attacking the brass components, soldered seams etc. Another reason is that, in order to secure the required low pressure using mercury, it would be necessary to employ a vapour trap of much more complicated design, which would in turn increase the flow resistance.

cussion of the necessary pumping speed), so that a difference in pressure occurs between the pump inlet and the outlet of the acceleration chamber. Let us denote the (constant) pressures at these points by p_0 and p respectively, and the volumes of gas flowing per second (the local pumping speeds) S_0 and S respectively (fig. 19). Then, since the mass

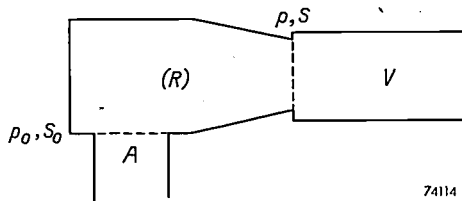


Fig. 19. In order to secure a pumping speed S at the acceleration chamber V against a pressure p , a greater pumping speed S_0 with lower p_0 is required at the pump inlet A , in view of the flow resistance (R) set up by intermediate elements in the pumping duct.

of the gas is the same at both points, $S_0 p_0 = Sp$; further, $p - p_0 = RpS$, where R is the flow resistance, as defined by this equation. From the two expressions t at once follows that:

$$\frac{1}{S} = \frac{1}{S_0} + R \dots \dots \dots (1)$$

It will be seen that S , the pumping speed at the chamber, can never be more than $1/R$, even if the speed S_0 of the pump itself were infinitely high. In the pump under discussion, $S_0 = 3500$ l/sec for air at a pressure $p_0 = 0.8 \times 10^{-4}$ mm Hg. After completion of the whole installation, a pumping speed of $S = 2200$ l/sec was measured at the acceleration chamber; hence the resistance is $R \approx 1.5 \times 10^{-7}$ sec/cm³, which is lower than was anticipated. (In view of the dimensions of the duct etc., we had estimated that R would be about 5×10^{-7} sec/cm³). Had we designed the pump for $S_0 = 7000$ l/sec, which would have been much more costly, the effective pumping speed would not have been increased to more than 3000 l/sec.

We cannot here enter too deeply into the working of the pump, but will mention only one or two points. The design is dependent to a considerable extent on the fact that it is mainly hydrogen, not air, that has to be pumped from the acceleration chamber. The vapour diffusion type of pump depends for its action on the incoming gases being entrained with the stream of vapour. Now, hydrogen is entrained at a higher rate and for this reason the pumping speed is appreciably higher than with air (theoretically by a factor of 3.78). At the same time, however, the gases are diffusing against the vapour stream at a certain speed, de-

pendent on the local pressure, thus obviously reducing the pumping speed, and this "back-diffusion" is more pronounced in the case of hydrogen than with air. To limit the effect of such "back-diffusion" as much as possible it was necessary to arrange for a very low backing pressure, a long stream of vapour and high vapour velocity v_d . In this regard careful dimensioning of the outlet slots and the choice of the lightest grade of oil were important; v_d is also increased according as the vapour pressure of the oil is raised (by greater heating of the boiler), but there is not much to be gained in this direction, as this results in too much dispersal of the vapour stream. The length of the stream is limited for the same reason, and the desired large "pumping distance" must therefore be achieved by means of stages in cascade.

On account of all these factors, the speed of our pump is roughly 4500 l/sec for hydrogen, this being about 1000 l/sec more than for air. Moreover, the flow resistance in the exhaust duct etc. (R in equation 1) is lower for hydrogen than for air, and this represents a further improvement in the pumping speed at the acceleration chamber.

The synchrocyclotron also includes another vacuum pump for evacuating the modulator drum (see I and II), but this pump does not have to meet such stringent requirements; a speed of the order of 500 l/sec is sufficient. None the less it was considered simpler and cheaper to make a pump identical with the one just described.

The pressure on the outer wall of the acceleration chamber (and also at other points) is monitored continuously by a Philips vacuum gauge (Penning type⁵⁾), an instrument that is well able to withstand sudden increases in pressure arising from possible faults occurring in the cyclotron. The action of this gauge depends on a gaseous glow discharge in a magnetic field; in the region in question the field of the cyclotron magnet — which is already available — is used. By making the electrodes of iron 3 mm thick it was found possible to reduce the flux density inside the discharge space to the required value of 0.04 Wb/m². Some of the pressure gauges used are fitted as warning devices; the electric discharge current flowing in the vacuum gauge, which is a measure of the pressure, is taken through an adjustable resistor and, at a given value, the voltage across this resistor ignites a mercury vapour tube, which in turn operates a warning signal by means of a relay. One such gauge incorporates an interlocking circuit that prevents

⁵⁾ See Philips tech. Rev. 2, 201-208, 1937; 11, 116-122, 1949.

the ion source from being switched in on the event of the pressure being too high, thus protecting the tungsten pin in the source from premature deterioration.

This brings us to the various safety devices, regarding which we shall say something in the following paragraph.

Controls and safety devices

To give an impression of the synchrocyclotron as a whole, after digressing into so much detail, let us now follow the procedure on commencement of each working period; we shall then describe some ancillary equipment.

First of all the two rotary vacuum pumps are started up, and the water supply for the high-vacuum pumps is turned on; voltage is then applied to the heaters of the pumps. The pumps acquire their normal working temperature in about $\frac{3}{4}$ hour, $\frac{1}{4}$ of an hour later the pump valve may be opened.

In the meantime the refrigerator for the vapour traps will have been started up, and the filament heater supply for the oscillator valves producing the Dee voltage, as well as that for the rectifier valves supplying anode voltage for the oscillator are then switched on; a delay of about 20 minutes is necessary before these valves can be put into operation. Cooling water for the oscillator anodes and the whole of the oscillator circuit including the Dees is next applied. If necessary, the tungsten pin in the ion source is renewed. The target, previously loaded with the required substance, is then introduced and connected to the cooling system. The rotary capacitor in the modulator is allowed to come up to its required speed, as well as the generator supplying the energizing current for the large magnet, and its field dynamo. The operation of opening the valves for the supply of cooling oil and water for the electromagnet unlocks a switch in the generator field circuit, which is next closed, and the current

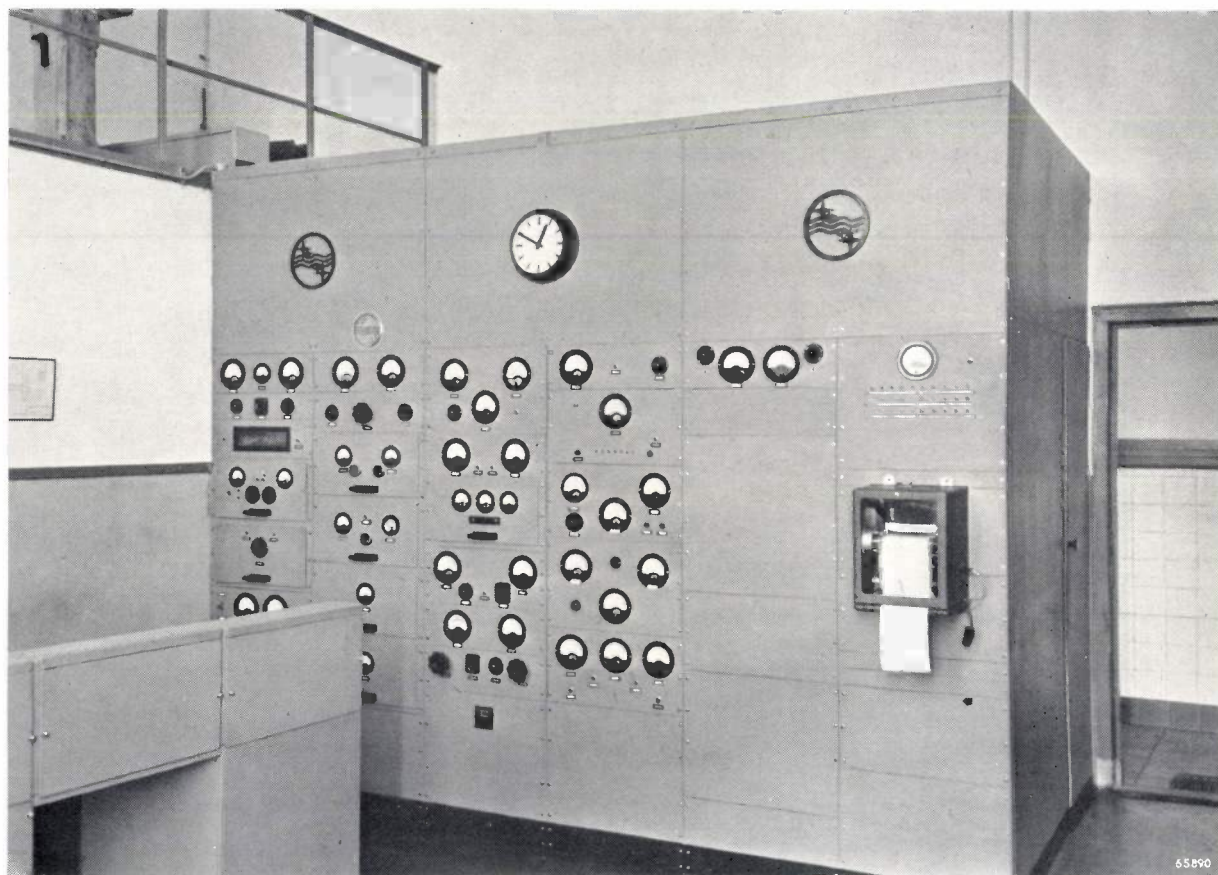


Fig. 20. The main panels in the control room of the synchrocyclotron, with switches, meters etc. for operating and monitoring the more important functions, viz. left to right: energizing of the magnet with stabilization and temperature monitoring device; anode voltage for the H.F. oscillator (with automatic cut-out); H.F. oscillator control (frequency measurement etc.); voltage and gas supply to the ion source and sundry other monitoring and control circuits such as on the speed of the modulator, the pumps, the continuous recording of the beam current and warning signals.

in the coil is adjusted to the correct value. The 500 V d.c. will in the meantime have been applied to the Dees, thus neutralizing possible discharge in the acceleration chamber (see II). At this stage voltage is applied to the anodes of the oscillator to supply the high frequency field for the Dees. Opening the water supplies to the ion source and target closes another interlock, so that, finally, the voltage can be applied to the ion source; before this is done, however, all personnel must have left the region of the cyclotron, as the beam of accelerated particles is formed and the target commences to produce radiation.

All further adjustments, as well as monitoring all the different elements of the equipment, must then be effected from the control room, which is isolated from the cyclotron by a 3.5 m thick wall of water (fig. 20). To commence with, the oscillator is tuned to resonance with the circulating particles, which is done by adjusting the trimming capacitor on the modulator (II, C", fig. 9) by means of a servomotor. The oscillator frequency remains reasonably constant and needs readjustment only once or twice during the day, this being done by hand. Other parameters which are subject to greater variation, are stabilized automatically.

Automatic stabilization is especially important for the magnet energizing current (the quotient of the average oscillator frequency and the magnetic flux density in the acceleration chamber must be kept constant within narrow limits⁶⁾). The method of stabilizing this current is illustrated in fig. 21. The energizing current passes through a high-stability resistor R (manganin wire resistor in an oil-bath maintained at constant temperature). The voltage across this resistor, in series with a stabilized reference voltage of opposite polarity, is applied to a circuit containing a mirror galvanometer. As long as the energizing current is at its normal value, the beam of light from the galvanometer remains at zero, but, should the current rise or fall by a given amount, the beam falls on one or the other of two photo-electric cells. The photo-current is amplified and applied to a servomotor which corrects the field (and consequently the voltage) of the field dynamo used with the main generator; thus, if the current exceeds the upper limit, the main generator field is accordingly reduced and vice versa. This "discontinuous" method of control has the advantage that it does not operate on variations in the current which are so small that they fall within the limits defined; there is therefore

⁶⁾ See W. de Groot, Cyclotron and synchrocyclotron, Philips tech. Rev. 12, 65-72, 1950.

little risk of "hunting" (continual adjustment backwards and forwards). The greatest variations in the current as stabilized in this manner are within 6×10^{-4} , and it follows from the magnetization curve of the steel (see III, fig. 3) that the flux density is then constant to within about 3×10^{-4} . This degree of stability is even higher than required.

Apart from the interlocking of certain controls already mentioned, the equipment is protected by numerous warning signals. In fig. 20 a series of lamps will be seen on the extreme right-hand panel, just above the instrument for recording the beam current. Should a fault occur in any of the components of the cyclotron, one of these lamps immediately lights as a warning. Faults involving any risk of damage immediately isolate the component concerned. For example, the pumps are automatically shut off in the event of the pressure in the system becoming dangerously high. Similarly, should one of the oscillator valves break down, the anode voltage is at once cut off; this is effected by means of the kind of repeater circuit employed in broadcasting transmitters: anode voltage is applied again automatically a few seconds later, to check whether the interruption is only transient; if flashover again occurs, the cut-out again opens, but after this,

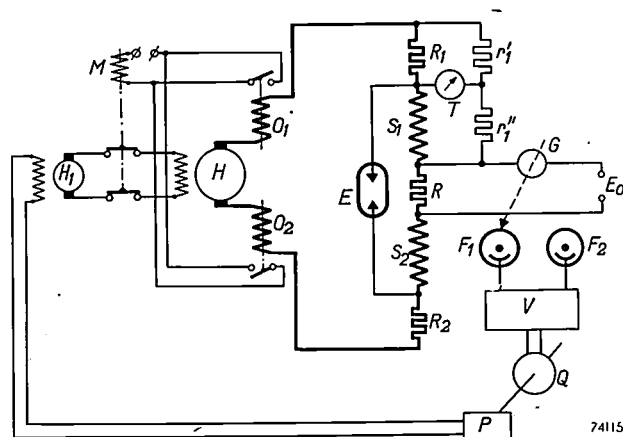


Fig. 21. The energizing circuit of the large electromagnet. S_1 and S_2 are the two large energizing coils, permanently connected to the generator H of the energizing current (500 V, approx. 175 A); H_1 field-exciter dynamo for the generator H ; O_1 , O_2 cut-outs operating on current overload, which break the circuit of H_1 through an electromagnetic switch M ; E rare gas fuse for protection of the coils. The energizing current is stabilized by means of a mirror galvanometer G which measures the difference between the voltage across a resistance R and a constant reference voltage E_0 . According to the position of the galvanometer light beam, two photo-electric cells F_1 , F_2 , by means of an amplifier valve V and a servomotor Q , control the rectifier P which excites the field dynamo H_1 . The galvanometer is protected against heavy current surges, particularly when the main circuit is closed and opened, by a rather complicated device not shown in the diagram. By means of a manganin strip R_1 and resistors r' , r'' , which together with the coil S_1 form a bridge circuit, and of the micro-ammeter T , the temperature of the coil S_1 is monitored continuously (a similar arrangement is provided for coil S_2).

the voltage will have one more "try"; only if the same result is obtained this time, is the anode voltage cut off finally and the fault has to be traced and rectified.

Mention should also be made of the special precautions taken to safeguard the energizing coils of the magnet. In view of the enormous amount of energy stored in the field of the magnet (theoretically 650 000 Wsec (see III) but in reality, in view of stray fields, even higher), care must be taken to ensure at all costs that the energizing circuit is never broken; the resultant high potential across the coils would otherwise have catastrophic consequences. In the first place, therefore, the coils are permanently connected to the generator (fig. 21). The only switch used is in the field circuit of the generator, as will already have become clear to the reader. Further there are no fuses between the coils and the generator. To protect the generator itself against possible short-circuiting of the energizing circuit, this circuit includes an overload-current relay for isolating the field dynamo from the main generator. The magnet coil circuit is thus not interrupted and the magnetic energy can leak away gradually through the generator. Connected in parallel with the coils, moreover, is a rare-gas fuse whose electrodes are held under pressure by a heavy spring; should the connection between the generator and the coil at any time be open-circuited, the rising voltage will ignite the gas-filled fuse. This in itself will provide a path for the discharge of the energy, but the resistance of this path is reduced in a matter of seconds because the glass discharge tube softens as a result of the heat developed and is telescoped by the spring. The electrodes then con-

stitute an effective short-circuit for any further rise in the voltage on the coils.

Summary. Continuing with the earlier articles on the oscillator, modulator and electromagnet of the Philips synchrocyclotron at Amsterdam, a description is now given of the other important components, viz. the acceleration chamber, the Dees, the ion source, target, vacuum pump and some of the controls and protective devices. The circular, evacuated acceleration chamber, nearly 2 metres in diameter, comprises 7 cm thick top and bottom plates of the steel of which the magnet is made; for the rest it is built up in situ, of conveniently small members, so that it can at all times be dismantled completely or in part. The vacuum joints are sealed with rubber strip and the total air leakage amounts to only 6 litres per week. At the gap between the Dees the height of the Dee is 20 cm, to allow of the collection of as many ions as possible to form the beam and to ensure that few will be lost as a result of collisions arising from the vertical oscillations. The ion source consists of a chamber to which heavy hydrogen is supplied at a pressure of about 0.1 mm Hg; this passes subsequently through a communicating duct 5 mm in width to the acceleration chamber. A gas discharge is maintained in the ion source which results in a highly concentrated jet of electrons being sent into the acceleration chamber. The way in which this empirically devised mechanism functions cannot yet be fully explained, but the very copious supply of ions thus obtained is largely responsible for the high yield of radio-active materials which the equipment produces. The most satisfactory position of the ion source between the Dees can be found while the cyclotron is working. The target, in which a good 500 W of power is dissipated, is thoroughly water-cooled. The temperature of the cooling-water, which may be measured and recorded continuously, is a direct measure of the beam current. Both the ion source and the target are introduced into — and withdrawn from — the acceleration chamber through traps. The acceleration chamber is permanently connected to a 4-stage high-vacuum pump having an un baffled pumping speed of 3500 litres of air per second or, as measured at the chamber, 2200 l/sec. A control room, which is isolated from the cyclotron by a wall of water 3.5 m thick, houses a number of controls and switches for monitoring and adjusting the working of all components during the operation. The energizing current for the magnet is automatically kept at a constant value by a stabilizing system using a mirror-galvanometer and two photo-electric cells. Possible faults in a number of the components of the cyclotron are indicated by means of signal lamps, or, in the event of serious breakdown, the component is automatically isolated. Errors in manipulation are prevented by a number of interlocking systems.

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

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.

2008: J. C. Francken: Opneembuizen voor televisie, III. Electronenoptische problemen in de beeldiconoscoop (T. Ned. Radiogenootschap **16**, 243-257, 1951, No. 5). (Television camera tubes, III. Electron-optical problems in the image iconoscope; in Dutch).

In the image-iconoscope two parts can be distinguished: the scanning section and the electron-image section. In the design of both parts

typical electron-optical problems are involved. With respect to the scanning section, means to obtain nearly uniform focus over the target are discussed. It appears to be advantageous to use a high-current density at the cathode. Therefore, use is made of the so-called L-cathode which also allows for accurate mounting of the gun parts. The electron-image section consists of the photocathode, an electron lens and the target. The elec-

tron lens is formed by co-axial electrostatic and magnetic fields. The electrostatic field exists between cathode and anode cylinder, the magnetic field is generated by a rather long coil. The properties of this system are discussed. Causes of unsharpness (chromatic aberration and curvature of the field) and geometric distortion (pin-cushion and S-distortion) are mentioned. All these errors can be reduced to such an extent that a satisfactory image can be obtained. A special feature of this type of electron lens is the possibility of varying the magnification of the image continuously without affect its sharpness and rotation. This can be done by using a coil split up into three separate sections. By varying the currents in these sections simultaneously, a continuous variation of the magnification in a range of 1:2 can be achieved (see Nos. 2006 and 2007).

2009 *: P. Schagen: On the image iconoscope, a television camera tube (Thesis, Amsterdam 1951, 89 pp, 40 figs.).

After a short historical introduction and a brief discussion (Ch. I) of modern television camera tubes, a simplified theoretical approach is given in Ch. II to the processes involved in building up the potential image on the target of the image iconoscope by the photo-electrons, and the way in which the video-signal is created when the potential image is removed by the scanning beam. This theory, based on some hypotheses concerning the behaviour of secondary electrons, leads to expressions for the potential of a target element as a function of the photo-electric current.

In Ch. III the methods are described with which the signal output and the secondary-emission coefficient of the target have been measured.

The theoretical results are compared with the measurements in Ch. IV and some conclusions are drawn with respect to the steps which can be taken to improve the characteristics of the tube. It appears that the resulting experiments confirm the theoretical predictions.

In Ch. V a method is described which considerably reduces spurious signals in the video-signal produced by the image iconoscope, and enables a line-by-line restoration of the black-level.

Finally, in Ch. VI, the performance of the image iconoscope is discussed in comparison with other camera tubes as regards the relation between image signal and photo current, effective time of illumination, spurious signals, simplicity of operation and focussing, spectral sensitivity, stability, black-level, variable magnification, depth of focus, signal-to-noise

ratio and sensitivity. Attention is drawn to the fact that sensitivities are to be compared by comparing the illumination levels needed for producing images of good quality with corresponding depth of focus and noise.

The conclusion is drawn that the Philips image iconoscope is a very good tube for studio work as well as for outdoor work in daylight. For illuminations below 1000 lux the quality of the image may be insufficient for broadcasting purposes. This might be improved by producing a tube with a larger photocathode.

2010: J. M. Stevels: La coloration du verre (Verres et Réfractaires, 5, 197-204, 1951, Aug.). (The colouring of glass; in French.)

The colouring of glasses is determined:

1) by the slope of the absorption curve, 2) the presence of atoms having a specific absorption, 3) by electronic transitions between cations. The location of the slope of the absorption curve is determined by the Madelung potential at the place of an oxygen ion. This potential itself depends on the location of the oxygen ion and on the charge of the network forming ions. Admitting as such the ions P^{5+} , Si^{4+} and B^{3+} , one finds six possible positions of the absorption curve.

The absorption caused by ions with specific absorption in the visible spectrum essentially depends on their location in the lattice (network former or network modifier).

Transitions between cations (network former or modifier) complicate the absorption problem. This phenomenon presents itself especially if one sort of ion occurs with two different valencies. The case Fe^{3+} , Fe^{2+} is representative and very instructive in this respect.

2011: E. J. W. Verwey: Ver-reikende krachten in colloïd-chemische en biochemische systemen (Hand. Nat. Geneesk. Congres 32, 4-14, 1951). (Long distance forces in colloidal and bio-chemical systems; in Dutch).

This lecture deals with the forces, attractive (London-Van der Waals) and repulsive (electric double layer) between colloidal particles (see these abstracts, No. 1769) and with the question of stability of colloids. The same considerations are tentatively applied to bio-colloids. In this case, beyond general repulsive and attractive forces, there are specific forces, dependent on the structure of the particles. The experiments of Rothen on interaction between antigen and antibody are mentioned as an example. Finally a number of phenomena occurring in hydrophilic colloids are discussed.

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 STEEL PICTURE-TUBE FOR TELEVISION RECEPTION

by J. de GIER, Th. HAGENBERG, H. J. MEERKAMP VAN EMBDEN,
J. A. M. SMELT and O. L. van STEENIS,

with an introduction by J. G. W. MULDER.

621.385.832:621.397.62

A milestone in the development of larger picture-tubes for direct viewing was reached with the tube which appeared on the market some years ago in America, and recently also in Europe, namely the tube with a conical portion made of metal. One of the many advantages of this is that the glass window may be made thinner, thereby permitting the use of drawn glass; this latter has a uniform thickness, which results in an improvement in image quality.

The first types of this new form of tube had a cone made from chrome-iron, a rather expensive material. It may therefore be considered a great step forward that success has been achieved in producing a type of glass which can be sealed to steel, and that at the same time a good solution has been found for the difficult corrosion problem involved.

A tube with a steel cone and a window made of the new glass is now coming into production. The authors describe the development of this tube. Their description is preceded by an introduction dealing with the history of metal discharge tubes.

INTRODUCTION

It has long been a cherished desire to make the walls of various discharge tubes as much as possible from metal and to use glass only where necessary for the insulation of the leads. New reasons for such a development are constantly arising: in transmitting valves it is attractive to allow the metal wall to serve as the anode, which can then be intensively cooled by water or by air; in the case of X-ray tubes operating in air, a metal wall is not only attractive for cooling purposes, but gives in addition good protection from undesired radiation and, if earthed, from the high-voltage danger; in industrial rectifiers metal tubes are more suitable than large more vulnerable glass tubes, etc. Similarly, in the case of picture-tubes for television reception — at least of the larger tubes for direct viewing — a metal wall has particular advantages, as will be seen in the following.

In order to realise this objective one must in the first place have available a metal and a glass which can be sealed together to give a vacuum-tight

joint; in addition the seal must be able to withstand the temperature changes and localised temperature differences which may arise during the operation of the tube. Less fundamental, yet in practice important, further requirements are that the metal shall not be too difficult to work, not too expensive, etc.

Acceptable solutions have been found in many cases. By the Housekeeper method, for example, in spite of considerable difference in expansion coefficient between glass and metal, a tight seal is obtained by the use of a ductile metal, such as copper, which is made very thin at the place of sealing. The metal can then, by yielding slightly, accommodate its dimensions to those of the glass. Although this seal forms a weak spot, from a mechanical viewpoint, it is nevertheless employed in some types of transmitting valves.

Mechanically stronger is the seal between chrome-iron and lead-glass — two materials which match each other very well as far as expan-

sion coefficient is concerned¹⁾. Use is made of chrome-iron in numerous tubes manufactured by Philips, either for the contact-pins or for the tube wall itself. This latter is the case in various transmitting valves, in X-ray tubes operating in air, and in some rectifiers and thyratrons (see the final reference quoted in note¹⁾).

There are certain objections to the use of chrome-iron for the wall, namely the poor workability and the costliness of this material. The same objections also obtain in the case of alloys such as fernico, which has a lower expansion coefficient, matching the harder glass types. In view of the constant demand for metal tubes, efforts have been made to overcome these disadvantages, by attempting to use as the metal a cheap and easily workable type of steel. Two lines of approach have been followed by Philips with this end in view. One approach has led to a particular seal construction, the other to a new type of glass, matching steel in expansion coefficient.

Compression seal

The first solution arose out of the fact that, in general, glass withstands compression better than tension (the ratio of the maximum permissible compression stress to the maximum permissible tensile stress can be as much as 20 : 1) and that it is in fact the tensile strain, with unsuitable construction methods, which is normally responsible for cracks. Because of this conviction we developed in 1940-41²⁾ a rectifier tube (fig. 1a), in which it was arranged that under normal circumstances the glass was under compression ("compression seal"). A seamless steel tube is shrunk around two lead-glass pressed forms (fig. 1b), and serves as the tube wall³⁾. At room-temperature the glass is under considerable compression. If the tube heats up during operation, it is true, this compression decreases, as the expansion coefficient of steel (13.0×10^{-6} per °C) is greater than that of lead-glass (9.5×10^{-6} per °C). Nevertheless there is a considerable useful range before the temperature reaches such a value that the compression strain changes over into one of tension, and the chance of cracking becomes significant.

1) B. van der Pol, Metal transmitting triodes for high power, *Sterkstroom I*, 265-266, 1923 (in Dutch); A. Bouwers, A new X-ray tube, *Physica* 4, 173-179, 1924 and A New Metal X-ray tube, *Fortschr. Röntgenstr.* 33, 575, 1925. A summary is given in an article by H. J. Meerkamp van Embden, *Joints between metal and glass*, *Philips tech. Rev.* 2, 306-312 1937.

2) For reasons outside the scope of this discussion, this tube was never brought on the market.

3) A procedure for making a compression seal between a steel pipe and pressed glass is described in Dutch patent application No. 102 638.

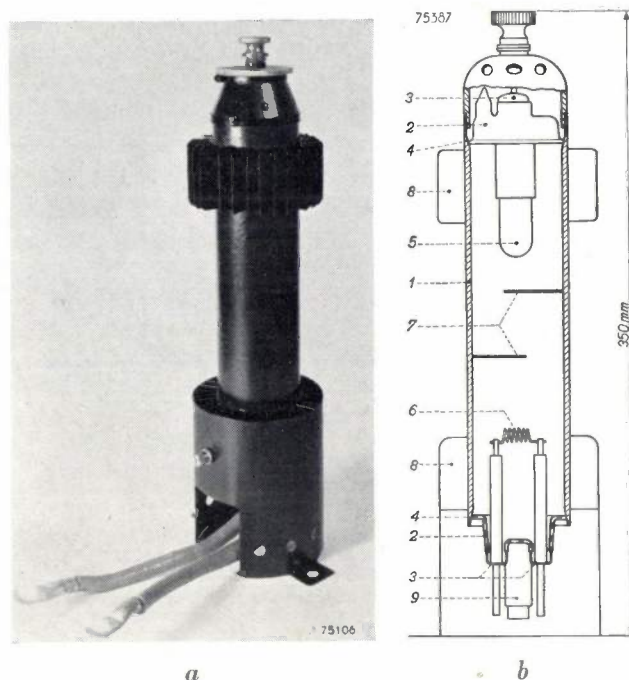


Fig. 1. Steel-walled rectifier tube with compression seals. a) photograph, b) cross-section of the valve made in 1940-1941. The tube wall consists of seamless steel tubing 1; the ends 2 are of pressed lead-glass, with chrome iron caps 3 for the current leads. At the seal 4 between steel and glass, the glass is under compression. 5 anode, 6 directly heated oxide-coated cathode, 7 partitions, which locally constrict the current path and thereby hinder arcing-back, 8 cooling fins, 9 mercury reservoir.

The tube contains mercury vapour and argon, and under a tension of some hundreds of volts is capable of delivering a current of 15 A average value with a peak value of 75 A.

Since the glass flows, more or less, it relieves itself of strain to a certain extent in time. Over a number of years we have followed the course of the compression strains in a series of samples of this compression seal. The results (fig. 2) show that with our construction the maintenance of compression strain is assured for many years.

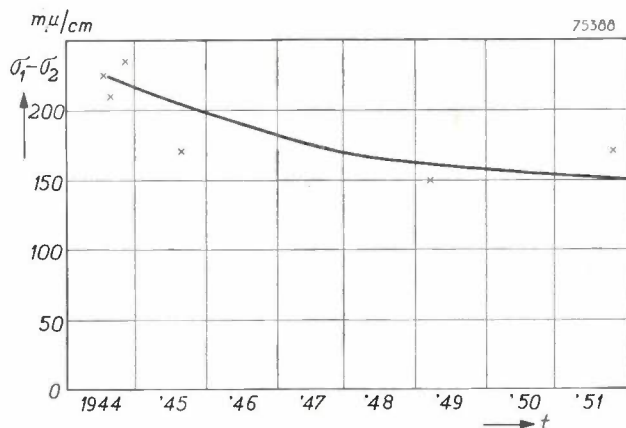


Fig. 2. Difference between two principal strains (tangential and radial), $\sigma_1 - \sigma_2$, in the glass of a compression seal, as a function of the time t over some years. $\sigma_1 - \sigma_2$ is expressed in $m\mu/cm$, the unit which is customarily employed for strain measurements in glass (see note¹⁾).

"Iron" glass

The other avenue followed was the search for a type of glass with approximately the same expansion coefficient as mild steel. The Philips Glass Factory in Eindhoven paid considerable attention to this project. This led in 1938 to a new type of glass (followed later by still more), which conformed to the requirement of sealability to iron and was

may arise at places where hot and cold parts of the tube border on each other. For this reason there is still a preference in the case of rectifiers and other heavily loaded valves for the combination chrome-iron/lead-glass (expansion coefficient $\alpha = 10.5 \times 10^{-6}$ and 9.5×10^{-6} per $^{\circ}\text{C}$ respectively) or still better, for fernico/hard glass ($\alpha = 4.6 \times 10^{-6}$ per $^{\circ}\text{C}$), rather than to the combination steel/"iron"glass

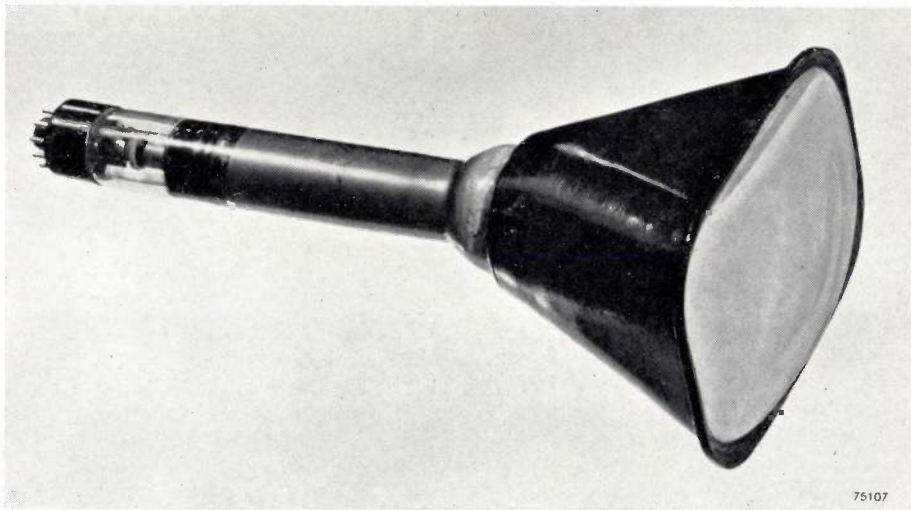


Fig. 3. Prototype picture tube made in 1949. Cone of mild steel, remaining parts of the bulb of (lead-containing) "iron" glass. Diagonal of the window approximately 15 cm.

therefore called "iron" glass. Rectifier tubes were constructed from this glass and mild steel in 1939.

Rectifiers, however, during operation or during warming-up and cooling-down, are subjected to large localised temperature differences. In such cases it is not sufficient that the glass and metal should have practically the same expansion coefficient, but the coefficient must in itself be fairly small; otherwise prohibitively large stresses

($\alpha = 13.0 \times 10^{-6}$ and 11.6×10^{-6} per $^{\circ}\text{C}$ respectively).

Conditions are different with picture-tubes, since in this case no large temperature differences arise. Fig. 3 is a photograph of a picture-tube manufactured in 1949, with a cone of mild steel and a window of "iron" glass.

The following article describes how the development thus introduced was successfully completed with the marketing of a modern steel picture-tube with a diagonal of 43 cm.

Picture-tubes for direct viewing

The first television receivers produced a picture of small dimensions according to present conceptions. In the years 1936-39 the standard tube in Europe had a diameter of about 22 cm; only in "de luxe" receivers was a tube of 31 cm employed.

After 1945 there arose a strong demand for larger images. These are less tiring for the eyes, and at the same time enable a larger company to view the picture. A change was made simultaneously to tubes with a more or less rectangular window

(fig. 3); in the case of large tubes, particularly, this permits a considerable saving in the dimensions of the cabinet.

There are two methods by which the demand for larger images may be met: the construction of larger tubes, and the use of an optical system which projects a considerable enlargement of the image produced on a small tube onto a screen (projection-viewing). Both methods are applied by Philips, but in this article we limit ourselves to the former ("direct view").

With increasing dimensions of (glass) picture-tubes, a number of objections become more and more serious:

1) In order to make the bulb able to withstand the pressure of the atmosphere, the tube window must either be made very thick or be given a strong curvature. In both cases the picture is deformed, so that the maximum usable picture dimensions do not increase proportionally to the diameter (or the diagonal) of the window. Even the

cone of chrome iron. In the course of this year, however, manufacture will go over to an externally similar tube with steel cone. This latter tube is the special subject of this article and is shown in *fig. 4*.

Special properties of metal picture-tubes

As we shall see later, a metal cone permits the use of a window which is fairly thin (5.5 mm) and which needs to be only slightly curved (radius of curvature approx. 75 cm) in spite of the force of



Fig. 4. Picture tube type MW 43-43 with rectangular window (diagonal 43 cm) and metal cone. During the course of 1953 the production will change over from chrome-iron to chromised steel as cone material.

best compromise between thickness and curvature is, generally speaking, not very satisfactory.

2) The larger the tube, the more serious the consequences of a possible implosion, against which safety precautions must be taken.

3) The cost and weight of the tube increase rapidly with increasing dimensions and, in the case of a medium sized tube, form a considerable proportion of the cost of the whole receiver.

It was therefore an interesting development when a tube of 41 cm diameter — which according to the ideas at that time was very large — came on the market in America in 1949, with a cone of chrome-iron. Philips make a similar tube (type MW 43-43) with a rectangular window of diagonal 43 cm and a

1.3 tons exerted by the atmosphere on the 43 cm diagonal window. These figures compare favourably with those of an all-glass tube of comparable dimensions: in such a tube the glass in the middle of the window is 7.5 mm thick and the radius of curvature is about 60 cm, whilst at the edge the glass is much thicker and more strongly curved (*fig. 5*). The whole area of the slightly and evenly curved window of the metal tube can be usefully employed without visible distortion at the corners of the picture.

The fact that the window of the metal tube is only slightly curved has further advantages. Firstly it gives a saving (albeit small) in the total length of the tube and hence in the necessary depth

of cabinet⁴). A second advantage is the following. The high brightness which is obtained with present-day direct view tubes, makes it unnecessary to completely darken the room in which one watches the television; for various reasons this is in itself undesirable. Thus, when one views in the evening, one or more lamps will usually be burning, so that the possibility arises that the viewers see distracting reflections in the tube window. In the case of the almost flat window of a metal picture-tube, it is not very difficult to arrange the lamps, the receiver and the viewers in relation to each other so that the latter no longer see the reflections. It is quite otherwise with the strongly and unequally curved window of a glass picture-tube; distracting reflections are then difficult to avoid.

The special form of the window of a glass tube, dictated by strength considerations, is obtained by pressing. It is, however, difficult to make pressed glass windows for these large tubes completely flawless. The surface possesses a certain roughness, which tends to spoil the sharpness of the image, and can be removed only by polishing.

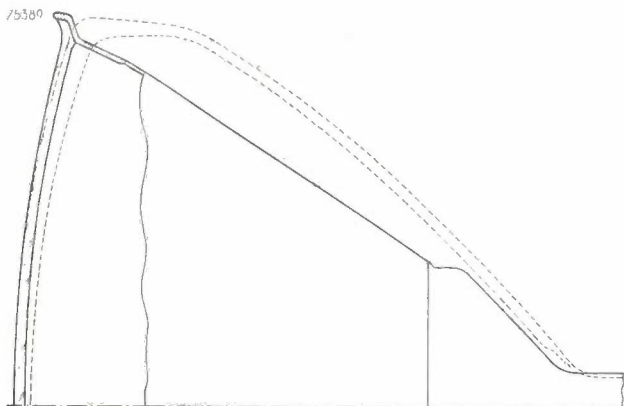


Fig. 5. Full lines: the steel-walled bulb of the tube MW 43-43 in cross-section. Broken lines: cross-section of a wholly glass tube with a window of the same size. In the former the window is less curved, thinner and of equal thickness all over.

This objection does not arise with metal tubes, since windows may be employed which are manufactured by a similar drawing procedure as for ordinary flat window glass and which possess a much smoother surface than windows of pressed glass.

The oldest fully automatic drawing process is that of Fourcault. In this an oval refractory stone — the “débiteuse” — having a slit, rests on the molten glass. The “débiteuse” is pressed downwards, thereby forcing the glass upwards through the slit. At the start of the operation the glass is taken up by a horizontal metal rod which is raised vertically.

The glass, which solidifies to a flat sheet, shows no tendency to constrict, and is thereafter continuously drawn upwards between rollers and repeatedly cut off at a given length.

Other well-known methods of drawing glass sheet are the Libbey-Owens and the Pittsburgh processes.

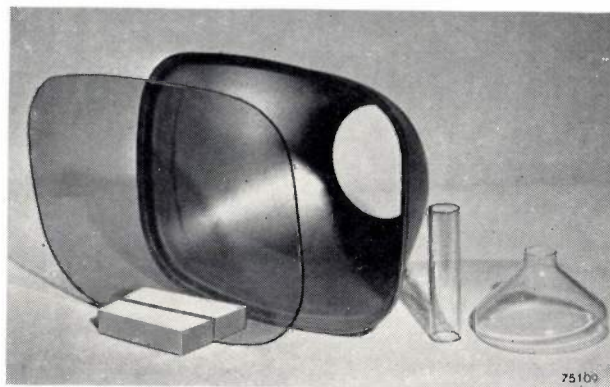


Fig. 6. Component parts of the bulb of the picture tube MW 43-43. From left to right: window of grey lead-free “iron” glass, cone of chromised steel, neck and dome of lead-containing “iron” glass.

Panels of the desired shape are cut from the flat sheet (fig. 6). These are pressed or “sagged” into a curved form and then sealed to the metal cone. As we shall see later, this is carried out in such a way that implosion of an exhausted bulb seldom, if ever, occurs.

The steel picture-tube

In the above-mentioned tube with chrome-iron cone, the chromium content was an important point of consideration. Chrome-iron with a content of 28% Cr was originally employed. Chromium is a rather expensive metal, however. In addition it is difficult to obtain chrome-iron of uniform composition; local faults are numerous and rejects are relatively high. A third disadvantage of chrome-iron with high Cr-content is its hardness, which renders it less suitable for the radical deformation from a flat sheet to a deep cone.

In our case, as also elsewhere, a change was therefore made to a material with an appreciably lower Cr-content (18%). The risk was thereby introduced of reaching an alloy range in which austenite may appear during the sealing-on of the glass, such an alloy being therefore really chrome-steel. The expansion coefficient of this austenitic chrome-steel is inacceptably high. The danger of austenite formation may, however, be excluded by the introduction of a small proportion of titanium or niobium. In this way a reliable and reasonably workable material⁵) is obtained, which is used on a

⁵) The metallurgical development work has been directed by Dr. Ing. E. M. H. Lips, of the Metallurgical and Plating Laboratory at Eindhoven.

⁴) This point forms the subject of an article to be published shortly in this Review.

large scale for the manufacture of metal picture-tubes.

Although an important saving in cost was achieved by the reduction of the chromium content from 28% to 18%, this alloy remains nevertheless appreciably higher in price than ordinary mild steel and many more or less special types of steel. Furthermore, most kinds of steel are softer than chrome-iron and thus require less working in the fabrication of a cone from sheet, so that simpler equipment is adequate. With these considerations in mind the development of metal picture-tubes could not yet be considered complete. In the use of steel for picture tubes, two main problems present themselves:

- 1) A glass matched to steel must be developed.
- 2) Precautions must be taken to protect the steel, both from chemical attack by substances with which it comes into contact during the manufacture of the tube and from rusting in the course of time. (Chrome-iron is fairly resistant, so that for this material the problem of protection raises no difficulty.)

We shall now discuss the glass and the corrosion questions in turn.

The glass

The "iron" glass developed earlier, as mentioned in the introduction, and used in the tube shown in fig. 3, is less suitable for the windows of picture-tubes for the following reasons:

- 1) It has a fairly low softening point, so that pumping of the tube at sufficiently high temperatures presents difficulties.
- 2) It is not easily drawn by the window-glass technique.
- 3) It contains a considerable proportion of a fairly costly material, namely lead.

These difficulties have now been overcome by the development in the Philips Glass Factory in Eindhoven of a new "iron" glass, which is completely lead-free⁶⁾. The principal characteristics of the old and the new "iron glass" are given in the table below.

The composition of the new glass is so chosen that the expansion coefficient of the steel and that of this glass have the same relationship as the expansion coefficient of chrome-iron and that of its appropriate glass. The implosion risk — to which we will refer later — is thus no greater than in chrome-iron tubes. The expansion coefficient of the

glass and that of the steel are regularly checked by sealing samples of these materials to a standard glass, in accordance with a fixed procedure. The strain at the boundary region, measured with the help of polarised light and a compensator⁷⁾, is a measure of the difference in expansion coefficients.

We see from the table that the temperatures at which the viscosity reaches certain values (in glass technology comparisons between glasses are always made on this basis), are 40 to 50 °C higher for lead-free glass than for lead-containing glass. This difference is large enough to make good degassing of the tubes possible.

The glass has a considerable chemical resistance. It may be compared in this respect with the glass used for chrome-iron picture-tubes.

When the tube is in use, the fluorescent screen and the cone have a potential of 12-14 kV with respect to earth. The specific resistance of the glass is so great that in spite of this high voltage on the inside of and around the window, no appreciable voltage appears on the outside.

Table. Some characteristics of a lead-containing and a lead-free "iron glass".

	Lead-containing	Lead-free
Expansion coefficient between 30° and 300 °C	11.6×10^{-6} per °C	11.8×10^{-6} per °C
Compression strain after being sealed to mild steel, in $m\mu/cm$ *)	350-400	350-400
Lowest annealing temperature **, (°C)	400	450
Highest annealing temperature ***, (°C)	425	470
Softening point †), (°C)	605	645
Temp. (°C) at which specific electr. resistance ††) = $3.33 \times 10^8 \Omega cm$	370	355
Specific electrical resistance (Ωcm), at 50 °C at 100 °C	3×10^{15} 7×10^{13}	1.5×10^{15} 7×10^{12}

*) See article referred to in note 7).

***) Temperature at which the viscosity reaches the value $10^{14.6}$ poise.

***) Temperature at which the viscosity reaches the value $10^{13.6}$ poise.

†) Temperature at which the viscosity reaches the value $10^{7.6}$ poise.

††) Measured as the temperature at which a rod of the glass, 50 mm long and of 5 mm diameter, passes a current of 1.18 μA at a potential difference of 100 V.

⁶⁾ J. Smelt, Le verre pour la télévision, Verres et réfractaires 6, 75-82, 1952 (No. 2).

⁷⁾ A. A. Padmos and J. de Vries, Stresses in glass and their measurement, Philips tech. Rev. 9, 277-284, 1947/1948.

By the use of certain additions the glass can be made grey, i.e. uniformly absorbent in the whole visible spectrum (see fig. 6). In compensation for the loss of light caused by this absorption, the picture loses less contrast due to light coming from outside. This can be explained as follows. To reach the eye of the viewer, the light falling on the fluorescent screen from outside must pass through the window twice, whilst the fluorescent light only passes through once. The stray light is thus attenuated more than the desired light, i.e. the contrast becomes greater as the window absorption increases. A favourable compromise is reached at 35% absorption; the picture is then still bright enough and sufficiently contrasty in a lighted room.

Apart from the window, the tube is composed of the following glass parts: the stem, on which the electron-gun is mounted, the neck, into which this stem is sealed, and the so-called dome, which forms the bridge between neck and cone. The window, cone, neck and dome are shown in fig. 6. A few words now follow about the glass used for the stem, neck and dome.

Even more than in the case of the window it is important that the glass of the dome and neck is of very large specific resistance, as it forms the insulator between the cone, which is at very high tension, and the conducting layer on the inside of the glass on the one hand and the earthed focusing and deflection coils on the other. Although the new, lead-free, "iron" glass is sufficiently insulating to serve for dome and neck, preference is given for these parts to the still better insulating lead-containing "iron" glass, with an eye also on the lower softening point of the latter. This simplifies the manufacture of the stem and also the sealing-in. The softer nature of this glass necessitates, naturally, that during pumping the parts mentioned must be somewhat less strongly heated than the window. This is no disadvantage, as the fluorescent screen is mainly responsible for the gas evolution.

The stem is also made of lead-containing "iron" glass. It is made ring-shaped, so that the distance between the lead-in wires is as large as possible.

Various materials have been tried as lead-in wires for the stem.

Pure iron wire ("Armco"-iron) may be employed provided that care is taken that it is not so exposed to oxidation that a thick, porous oxide-layer is formed. It is advantageous to pre-oxidise in a special atmosphere: in carbon dioxide and steam a dense, hard oxide is formed. To prevent over-oxidation, however, this must be immediately covered with an enamel layer, whereafter the wire must be very carefully handled.

Even less satisfactory are copper-clad iron wire, which develops small cracks around the wire, and chrome-plated iron wire, which displays irregular diffusion phenomena after heat-treatment.

Finally, chromised steel wire and dumet wire (nickel-steel wire in a copper sheath) were found to be very reliable. Chromised material is discussed further in the following paragraph.

Protection against corrosion

During the sealing-in of the window, when localised temperatures between 1050° and 1100 °C are reached, a thick oxide layer would be formed on the steel if the latter was not protected. This skin, loose in structure and porous, would make the the vacuum-tightness of the seal extremely doubtful.

Further risks of attack arise during the introduction of the fluorescent layer, which is allowed to settle from a weakly alkaline liquid⁸⁾. By this treatment, and also during the course of a few days, by the action of the air, loose particles of rust may be produced which, after dropping from the metal, can appear on the screen as dark spots or, still worse, get on to the cathode and impair the emission.

Lengthy trials were necessary to find a satisfactory way of overcoming the risk of corrosion. The following methods have been investigated: enamelling, nickel plating, passivating and chromising. Of these, chromising has led to the most reliable solution. The chromising process⁹⁾ is widely used in Germany for making all types of steel articles chemically more resistant. The process results in the formation of a thin layer of chrome-iron on the surface of the steel by diffusion of chromium, thereby supplying the desired corrosion-protection.

The chromising is carried out as follows. The articles to be treated — in our case steel sheets 3 mm thick — are heated with ferro-chrome of very high chrome-content in a closed oven, to about 1000 °C in a hydrogen atmosphere. From time to time a small amount of gaseous HCl is blown through the oven. This forms with ferro-chrome the volatile chromium chloride. The chromium atoms of the volatile compound now undergo an

⁸⁾ See the illustrations in Philips tech. Rev. 9, 339, 1947/1948 and 10, 306, 1949.

⁹⁾ A description of this process, which appeared in Germany shortly before 1939, is given by R. Becker, K. Daevs and F. Steinberg, Oberflächenbehandlung von Stahl durch Chromdiffusion, Stahl u. Eisen 61, p. 289, 1941; Metallwirtschaft, 20, p. 217, 1941; and Z. Phys. Chem. A. 187, p. 354, 1940. See also R. L. Samuel and H. A. Lockington, The protection of metal surfaces by chromium diffusion, Metal Treatment, 18, Aug.-Dec. 1951 and 19, Jan.-Febr. 1952.

exchange process with the iron atoms of the steel, which have practically the same size. In this way a layer of chrome-iron is formed on the surface; the layer has a thickness of from 70 to 150 μ , depending on the temperature and time of treatment. The chromium content is about 30% at the outside, decreasing towards the inside. The layer is very intimately bonded to the parent material. The latter can be etched away, however, and *fig. 7* shows the chrome-iron skins which remain after the steel has been etched away. This test demonstrates at the same time that chrome-iron is more corrosion-resistant than steel.

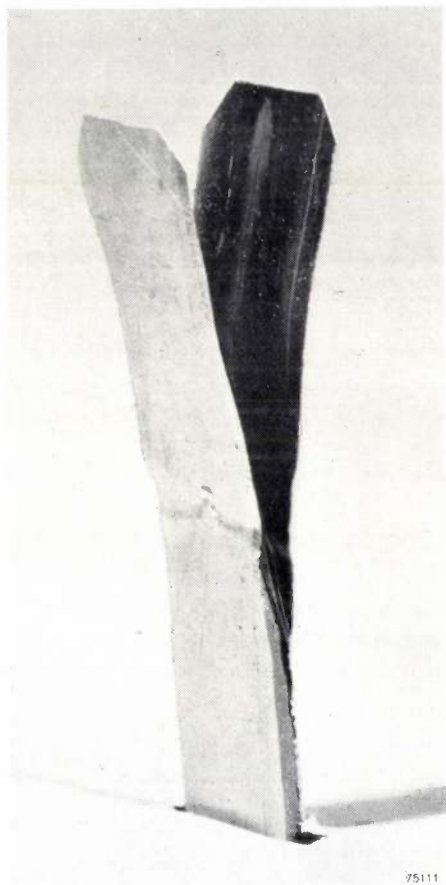


Fig. 7. Demonstration test of the corrosion resistance of chrome iron produced by chromising steel sheet. A portion of the steel is etched away, whilst the two chrome-iron skins remain unattacked.

The chromised sheets are easily fabricated into a cone, provided that certain precautions are taken and that the steel composition conforms to particular requirements. Thanks to the high Cr-content at the surface, the sealability to glass is very good. Since the chrome-iron layer is very thin, the chromised sheet has an expansion coefficient which hardly differs from that of unclad steel. Nevertheless the layer is sufficiently resistant to repeated withstand the treatment which a bulb undergoes

during manufacture; the cone from a rejected bulb can thus be used again, without having to be re-chromised. To protect the exterior of the cone over a long period against rusting, and at the same time to improve the outward appearance, it is lacquered at the end of the manufacturing process.

Chromised steel wire is used for the lead-in wires of the stem. This wire is 0.5 mm diameter, the chrome-iron layer being 15 to 20 μ thick.

Manufacture of bulbs with steel cone

Fig. 8 shows the cone in various stages of manufacture. The forming takes place in two steps. The 3 mm thick plate, cut to the desired size, is first spun into a round cone the height of which is about $\frac{5}{6}$ of that of the final product. The thickness is reduced by this treatment by about one half. The intermediate product is now placed under a heavy press. A punch and die are used to provide the cone at the wide end with an opening of the correct shape and size ("rectangular", with rounded corners and slightly curved sides). At the same time the rim acquires the profile illustrated in *fig. 5*. In this second treatment, which is more a folding than a drawing process, the material undergoes no reduction in thickness. The rim thus has the same thickness as the original sheet (3 mm), as is required to withstand the heavy mechanical stress present in the rim after exhausting.

After the bottom is punched out and the superfluous portions of the rim are removed, the cone is ready for the sealing-in of the glass parts. The dome is sealed in first (already joined to the neck, into which the stem with the gun will later be sealed), and then the window. In mass production, machines are employed for the processes involved in the sealing-in of the window. Twelve rotating heads are mounted on a slowly revolving table as in the sealing of lamps and radio valves. A cone (with dome and neck) and a window are inserted regularly as a position comes free. Several large flames of gas and air pre-heat these parts successively, and in one of the following positions of the revolving table the actual sealing-in is effected by means of a large number of smaller flames of gas and oxygen.

The sealing-in is completed before the bulb has made a complete revolution of the table. The upper part of the bulb then passes into the so-called equalising oven, the purpose of which is to bring the rim of the cone and the window to the same temperature. This temperature is so chosen that the glass remains somewhat soft.

The bulb leaves the oven after about seven minutes and is then allowed to cool in the air. This

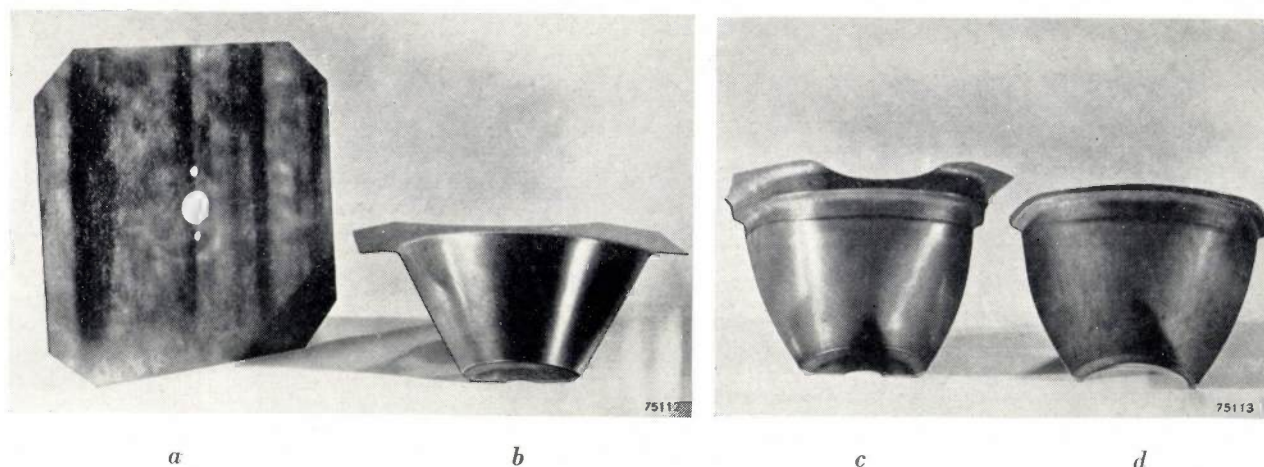


Fig. 8. Forming of a steel cone. *a*) 3 mm thick sheet of chromised steel, cut to the correct size. *b*) This plate is formed by spinning into a circular truncated cone. *c*) The cone has been given the desired shape, with the aid of a punch and die (the cross-section has a 'rectangular' shape with rounded corners and somewhat curved sides), whilst at the same time the rim has been given the correct profile. *d*) Superfluous portions of the edge and bottom have been removed.

differs from the commonly used cooling procedure for most all-glass objects; these remain for a long time in annealing ovens, in which they are brought to room temperature very gradually. There the intention is to achieve a strain-free end-product. In our case, on the other hand, it is in fact favourable if a certain amount of strain remains in the cooled glass, and in particular a compression-strain (see the Introduction).

Strain in the glass

The stress condition which will result in the present case depends on the expansion coefficients a_m and a_g of the metal and the glass respectively, the temperature of the equalising oven and the temperature T_s at which the glass begins to become elastic (i.e. at which stresses begin to develop in the previously soft glass).

As soon as the bulb leaves the oven, the cone and window begin to decrease in temperature; because it is steel, the cone cools off more quickly. When the glass reaches the temperature T_s , therefore, the metal is already at a lower temperature. If $a_m = a_g$, the glass would certainly reach a condition of tension with further cooling, as the circumference of the rim will contract further by a smaller amount than the circumference of the window. Actually, however, $a_m > a_g$ ($a_m = 13.0 \times 10^{-6}$ per $^{\circ}\text{C}$, $a_g = 11.8 \times 10^{-6}$ per $^{\circ}\text{C}$). The possibility thus arises in principle that the glass is under compression in the final condition. The proviso in this connection is that, at the moment at which the glass reaches the temperature T_s , the temperature of the metal shall not have decreased too far below T_s . This stipulation is easier to meet when the temperature

of the equalising oven is not chosen too high above T_s . By regulation of the oven temperature it is thus possible to bring the glass into the desired compression condition.

During the cooling of the glass another effect appears. When the hot window leaves the equalising oven, the outer layers of the glass cool more quickly than the inner ones. The latter will still be plastic at the moment at which the outside begins to become elastic. With further cooling, therefore, the outsides come under compressive forces. This effect ("toughening") gives an extra strength to the glass and is therefore often applied to the windows of motor-cars. The effect is proportionately stronger as the temperature of the equalising oven is made higher, but as we have seen above, the pressure exerted by the cone on the window is then smaller. One is thus forced to a compromise in the choice of equalisation temperature.

To obtain an impression of the strength of the bulb, it can be sealed off at the neck-end and subjected in a pressure vessel to a gradually increasing over-pressure. The bulb window will become flatter to a certain extent and hence attempt to increase its circumference, i.e. the tangential compression-strain originally present becomes smaller and changes over at a certain moment to a tension. With an over-pressure of 3 to 3.5 atm. the window fractures (fig. 9). The radial cracks show that the glass has failed due to tangentially operating tensile stresses. The metal is still intact, thanks to its greater tensile strength (~ 2100 kg/cm 2 , compared with that of the glass 300 to 700 kg/cm 2). It is for this reason that the bulb has not imploded, i.e. the glass splinters remain in place. Sometimes

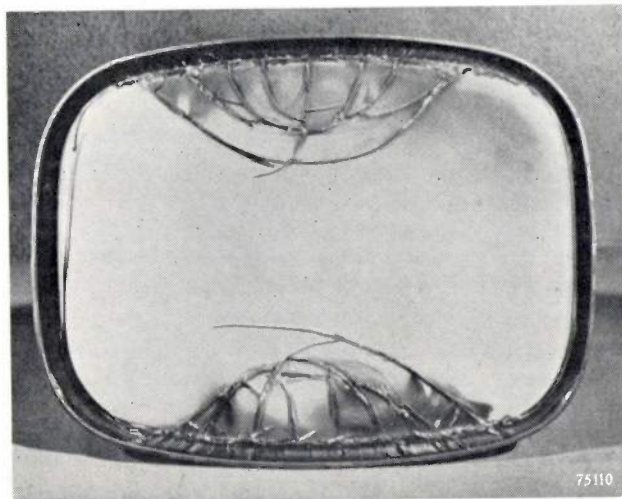


Fig. 9. A sealed-off bulb is exposed to excess pressure, which is increased until the window fractures. The glass does not implode; the splinters, on the contrary, remain in place.

they are pressed so strongly together that a tube with a cracked window can still be fully evacuated.

Investigation of the strain pattern

From the above it will be clear that it is very important to have the correct strain conditions in the window. A check on this is thus indispensable. This is carried out by tests on random samples. A small percentage of the manufactured bulbs are exposed to an over-pressure, as described. A larger percentage is subjected to a non-destructive examination with the help of polarised light. (See, for example, the article referred to in note ⁷). This method depends on the fact that mechanical stresses cause strain in the glass and make it double-refracting, and that the extent of the double-refraction is proportional to the applied stresses, or, more precisely, to the difference of the principal stresses. The proportionality-factor depends on the type of glass. The double-refraction can be detected by a change in colour of the transmitted light. The test consists in partially evacuating a bulb. The strains caused by the excess pressure of the atmosphere superimpose themselves on those already present. Thus it is possible to determine the under-pressure Δp_0 in the bulb at which the strains at the edge of the window exactly compensate each other; the colour of the edge area then changes over. The criterion for the assessment is that Δp_0 must be between two empirically determined limits.

Further description of the strain examination

A polarised beam of light — obtained by use of a "Polaroid" filter — passes through the bulb; it enters through the window and goes out through the dome. It then passes through a

so-called red plate¹⁰) and a second "Polaroid" filter. This latter transmits only light whose vibration direction lies perpendicular to that of the former. The beam passing through the second filter is observed.

The principal stresses in the glass vary through the thickness at the point considered, as a consequence of the curvature of the window. In fig. 10a this is depicted for one of the principal stresses, in this case for the tangentially directed principal stress σ_1 . For the radially directed principal stress σ_2 (perpendicular to the plane of the illustration) a similar condition arises; the third principal stress, perpendicular to σ_1 and σ_2 , is negligibly small. The variation of σ_1 in the thickness of the window can be represented by the superposition of a uniform compression stress (or it might be a tensile stress), σ_{10} (fig. 10b), and a symmetrically distributed bending stress (fig. 10c). The same applies to the radial stress σ_2 . The bending stress does not contribute to the optical effect since the double refraction caused by the positive bending stresses is exactly compensated by an opposing double refraction caused by the negative bending stresses. The resultant double refraction at a point in the window is thus a measure only of $\sigma_{10} - \sigma_{20}$ at that place, and this applies also in good approximation to the total double refraction (of window and dome together), since the glass of the dome is practically strain-free and thus causes no double refraction worth mentioning. It should be noted that the method gives no definite information about σ_{10} or σ_{20} independently, but only about the difference between these two principal stresses.

This stress difference $\sigma_{10} - \sigma_{20}$, depends on the distance r of the point under consideration from the centre of the window. Let us assume that the window is circular, as was the case with older tubes. It follows then from the symmetry that in the centre σ_{10} is equal to σ_{20} . No double refraction appears, therefore, i.e. no colour change is to be seen in the centre of the window. Outside of the centre colour variation is observed, in proportion to the magnitude of the local strain-difference.

The strain pattern in the window of an evacuated bulb is due to the superposition of the straining imposed during the sealing of the window and its subsequent heat-treatment and of the strain which arises from the over-pressure of the outside air. These stresses can be separately calculated for the rotationally symmetrical case described here¹¹); for a rectangular tube this is not possible. Thus the curve l in fig. 11 gives the calculated variation of $\sigma_{10} - \sigma_{20}$ as a function of r , in so

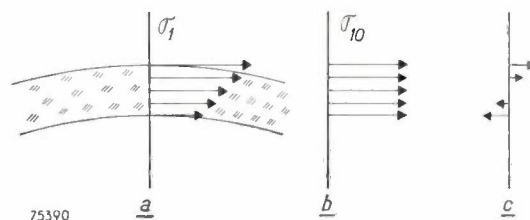


Fig. 10. a) Cross-section of the window of a metal picture tube, showing the variation of the tangentially directed principal stress σ_1 along the perpendicular at a given point. This variation can be represented by the superposition of an evenly divided compression (or tension) stress σ_{10} (see b) and a symmetrically divided bending stress (see c).

¹⁰) The function of the red plate is as follows. Without this plate the interference colours produced by double refraction in the glass are grey-blue in tint and therefore difficult to interpret. The red plate gives an appreciable, constant, double refraction corresponding to the interference colour "red of the first order". Changes in this colour as a consequence of double refraction in the object examined are easy to interpret.

¹¹) The calculations were carried out as suggested by Dr. Ir. J. A. Haringx of the Philips Research Laboratories, Eindhoven.

far as the stresses are caused by the difference between the expansion coefficient of the metal (α_m) and that of the glass (α_g). The curves 2 give the calculated values of $\sigma_{10}-\sigma_{20}$ in so far as these stresses are caused by the difference in air pressure; each curve corresponds to a certain pressure difference Δp . Superposing the curve 1 and one of the curves 2 gives the curve 3, which shows the total difference in principal stress at a certain pressure difference Δp . The magnitude of Δp can be so chosen in the test described that 3 passes through zero when r is equal to the radius R of the window, as represented in fig. 11. In this condition the glass shows no double refraction at the edge.

In this way we arrive at an empirical check on whether the sealed-in windows have the correct strain pattern. If air is pumped out of the tube until the colour of the edge area is

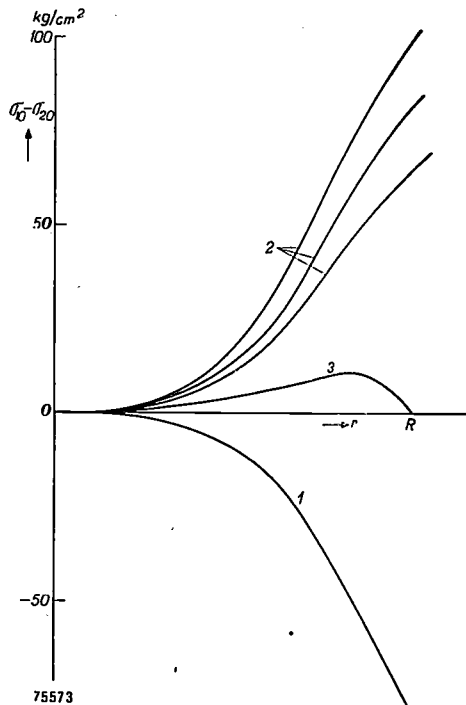


Fig. 11. The variation of $\sigma_{10}-\sigma_{20}$, calculated from the expansion coefficients α_m and α_g (curve 1), and the calculated values of $\sigma_{10}-\sigma_{20}$ resulting from various differential air pressures Δp , (curves 2), as functions of the distance r from the centre of a circular window (radius R). Curve 3 is obtained by superposition of curve 1 and one of the curves 2, so selected that 3 goes through zero at $r = R$.

Summary. One of the ways in which the demand for larger television images has been met, is the construction of larger picture-tubes for direct viewing. An important step in this development was the change-over from all-glass tubes to tubes with a metal conical portion. The window can then be drawn according to window-glass technique. A much more even thickness can be obtained by this procedure than by glass pressing, which is applied for the production of the thicker, stronger curved window of all-glass tubes. This results in an improvement in picture quality. Thanks to the reduced curvature, the whole window surface is usefully employed and reflections from lamps in the room are easily avoided.

In the most recent development the cone metal is made of mild steel, with which useful experience in similar applications had already been gained. It is considerably cheaper and more easily worked than the chrome-iron used hitherto. A new type of glass had to be developed, however, which, while having an expansion coefficient near to that of steel, had also a sufficiently high softening point (with a view to

seen to change over, then that curve has been chosen from the group of curves 2 which the curve 1 compensates, at least at the edge. The pressure difference Δp_0 at which this occurs is a measure of $\sigma_{10}-\sigma_{20}$ at the edge. This difference in pressure

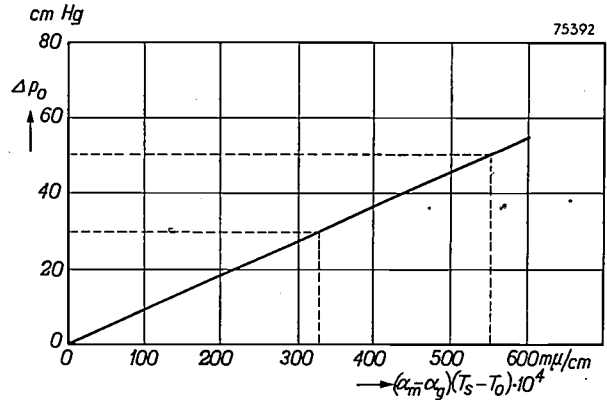


Fig. 12. The excess pressure Δp_0 which is necessary to make $\sigma_{10} = \sigma_{20}$ at the edge of the window, is a linear function of the difference in circumference $(\alpha_m - \alpha_g)(T_s - T_0)$, which the cone rim and the window would show if they were cooled down separately from T_s to T_0 .

is proportional to the amount of pre-straining in the window, i.e. to the relative difference in linear dimensions that the window and the cone rim would show if they had been allowed to cool down separately from the softening temperature T_s of the glass to the room temperature T_0 . The pressure difference Δp_0 is thus proportional to $(\alpha_m - \alpha_g)(T_s - T_0)$. This linear relationship¹²⁾ is shown in fig. 12, as calculated for circular tubes.

The calculation cannot be carried out for rectangular tubes, but a similar relationship holds and the same test can be made. In this case, the test is applied near the middle of the longest edge, this being the weakest spot. For rectangular, as for round tubes, the criterion is that Δp_0 must be between certain empirically established limits.

¹²⁾ As described on p. 282 of the article referred to in note ⁷⁾, the expansion coefficient of a material can be very accurately measured by means of a strain examination (with the help of polarised light) after sealing to a plate of a standard material.

The amount of the observed double refraction per cm thickness of the material is expressed in millimicrons wavelength difference in the light rays. For this reason the abscissae of fig. 12 and the ordinates of fig. 2 are provided with a scale in μ/cm .

the de-gassing of the tube.) Success has been achieved in making such a glass (lowest and highest annealing temperatures 450 and 470 °C respectively, softening point 645 °C). This glass is lead-free, which is favourable as far as price is concerned. Furthermore, it can be made grey to reduce loss of contrast caused by light coming from outside. To protect the steel from corrosion during fabrication of the tube, the chromising process was finally adopted, in which a layer of chrome-iron is formed on the surface. Chromised steel is also used for the lead-in wires of the stem. The sealing of the window into the cone and the subsequent cooling, are carried out in such a manner that in the final condition of the cone the edge of the glass is put under pressure and at the same time compression strain is set up in the outermost layers of the glass ("toughening"). The window owes its great strength, in spite of its small thickness, to this fact. The risk of implosion is appreciably smaller than with glass tubes of the same dimensions. The achievement of the correct strain pattern in the glass is checked with the aid of polarised light.

MECHANISM OF THE SYNCHRONIZATION OF LC-OSCILLATORS

by J. VAN SLOOTEN.

621.396.018

Electric oscillators whose frequency is governed by an inductance and a capacitance show changes in phase and frequency when influenced by an interference, the best known phenomenon being synchronization, where a periodic interference of sufficient strength causes the oscillator to vibrate in its own frequency. If the interference is too weak to synchronize the oscillator, a beat phenomenon arises in combination with a small variation of the mean frequency. Assuming the oscillator voltage to be practically sinusoidal, the author proceeds to deal with these phenomena with a view to giving a clear picture of their mechanism. With the method followed it is also possible to investigate the frequency fluctuations under the influence of a non-periodic disturbance, for instance the thermal noise voltage of the oscillator circuit.

Introduction

In radio engineering use is made, on a large scale, of electric oscillators which produce an approximately sinusoidal voltage. Mostly these consist of an oscillatory circuit formed by an inductance and a capacitance in combination with a triode feeding the circuit with a current which is in phase with the alternating voltage across the circuit. In a certain sense the triode then functions as a negative resistance shunted across the oscillator circuit. As a rule the amplitude is limited by a shifting of the working point towards a comparatively high negative grid bias as the amplitude increases. This takes place automatically by the rectifying action of the diode formed by the grid and cathode. The properties of these technical oscillators have already been discussed at length in this journal ¹⁾, where it was made clear that the manner in which the amplitude is limited influences greatly those properties of the oscillator which are related to the mean amplitude, as for instance the stability of operation. There are, however, also properties of the oscillator which, in the first instance, are independent of the amplitude-limiting mechanism. Among these is the susceptibility of the phase and frequency to extraneous disturbances, an example of which is the known property of an LC oscillator ²⁾ being synchronized by an external periodic signal of sufficient strength and of a frequency differing from the natural frequency of the oscillator.

About this synchronization phenomenon a number of papers has already been published, almost entirely based upon the original investigations of Möller, Appleton and, particularly, of Van der

Pol ³⁾. These publications are characterized by very general mathematical reasoning, so that also those oscillators whose damping and regeneration are so great as to cause the mode of oscillation to deviate greatly from a sine curve (relaxation oscillations) have been included in the investigations. When afterwards introducing into the results the assumption that the mode of oscillation is practically sinusoidal, it is found that there is a very simple condition for synchronization ⁴⁾. A disadvantage attaching to this method is that the mechanism of the synchronizing process is not made clear. When, however, one starts right from the beginning by assuming the mode of oscillation to be practically sinusoidal, it appears that an entirely different method of treating the problem is possible, a method which is not only simple but also has the advantage that no assumptions need be made about the way in which the amplitude of oscillation is limited. From the following description of this method it will be seen that in this way it is also possible to give an elementary treatment of the beat phenomenon occurring when the interfering signal is too weak to obtain synchronism. It is then only a step farther to investigate the irregular frequency fluctuations arising in the case of an arbitrary disturbance, as for instance a source of noise in the oscillator itself.

Phase variation under the influence of one single short pulse

Fig. 1 represents a common type of oscillator, where the interfering signal is assumed to act as a

¹⁾ J. van Slooten, Philips tech. Rev. 7, 40-45 and 171-177, 1942.

²⁾ For the sake of simplicity the term LC oscillator is used for an oscillator whose frequency is governed by an inductance L and a capacitance C .

³⁾ H. G. Möller, Jahrbuch drahtl. Tel. 17, 256-287, 1921.
E. V. Appleton, Proc. Cambr. Phil. Soc. 21, 231-248, 1922.
B. van der Pol, Phil. Mag. 3, 65-80, 1927.
B. van der Pol, Proc. Inst. Rad. Engrs. 22, 1051-1086, 1934.

⁴⁾ G. Francini, Alta Frequenza 18, 125-133, 1949.

source of current shunted across the LC circuit, which is represented by the pentode at the left, and the interfering voltage s applied to its control grid.

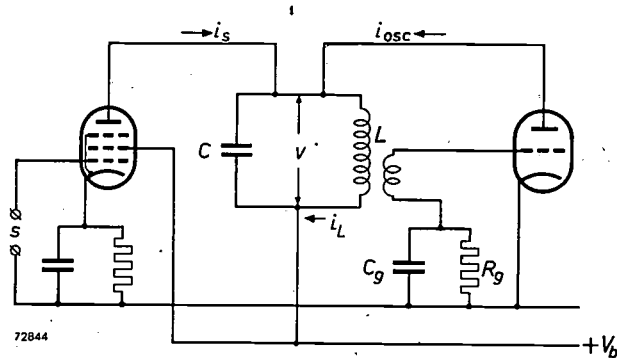


Fig. 1. A conventional oscillator circuit. The disturbing voltage s acts upon the grid of the pentode at the left.

The amplitude of the oscillation is limited by means of the grid capacitor C_g and the grid resistor R_g , but in the following considerations the nature of the limiting mechanism does not play any part, so that some other method of limiting, for instance by the curvature of the i_a-v_g characteristic, could equally well be assumed.

The line of reasoning is that the oscillator is to be regarded as a freely oscillating LC circuit whose amplitude is maintained at a constant level by the oscillator valve feeding the circuit with a current which is in phase with the circuit voltage v . Thus that valve influences the amplitude but not the phase of the voltage across the oscillatory circuit. In accordance with this supposition is the result, which will presently be found, that a disturbing pulse, coinciding with a voltage maximum or minimum of the circuit voltage v , does not affect the phase of the oscillation.

For the purpose of this investigation the effect is considered of a short pulse fed from the pentode in fig. 1 to the circuit at an instant separated from a preceding voltage maximum by a phase angle φ . This is represented in fig. 2.

Let the charge $i_s \Delta t$ given to the circuit by the interfering pulse be :

$$q_t = i_s \Delta t. \dots \dots \dots (1)$$

The current i_L through the coil being continuous, the interfering pulse causes a momentary rise in voltage Δv at the capacitor C , given by

$$\Delta v = \frac{q_t}{C}. \dots \dots \dots (2)$$

We shall now make the capacitor voltage v' and the coil current i_L' , immediately following upon

the pulse, correspond to that of an undisturbed oscillation, the phase and amplitude of the voltage maximum of which differ from those previous to the pulse. Hence:

$$\left. \begin{aligned} V_{\max} \cos \varphi + \Delta v &= V'_{\max} \cos \varphi', \\ I_{L\max} \sin \varphi &= I'_{L\max} \sin \varphi'. \end{aligned} \right\} \dots \dots (3)$$

It is, moreover, self-evident that

$$\frac{V'_{\max}}{I'_{L\max}} = \frac{V_{\max}}{I_{L\max}}.$$

Let us now put:

$$\left. \begin{aligned} V'_{\max} &= V_{\max} + \Delta V_{\max}, \\ I'_{L\max} &= I_{L\max} + \Delta I_{L\max}, \\ \varphi' &= \varphi + \Delta \varphi. \end{aligned} \right\} \dots \dots (4)$$

Since $\Delta \varphi$ is a small angle, we may write: $\cos \Delta \varphi = 1$ and $\sin \Delta \varphi = \Delta \varphi$. Employing a known trigonometric relation, we may then introduce in (3):

$$\left. \begin{aligned} \cos \varphi' &= \cos \varphi - \Delta \varphi \sin \varphi, \\ \sin \varphi' &= \sin \varphi + \Delta \varphi \cos \varphi. \end{aligned} \right\} \dots \dots (5)$$

Ignoring the terms containing the products $\Delta V_{\max} \Delta \varphi$ and $\Delta I_{L\max} \Delta \varphi$, it follows from (3), (4) and (5) that:

$$\left. \begin{aligned} \Delta V_{\max} \cos \varphi - V_{\max} \Delta \varphi \sin \varphi &= \Delta v, \\ \Delta I_{L\max} \sin \varphi + I_{L\max} \Delta \varphi \cos \varphi &= 0. \end{aligned} \right\} (6)$$

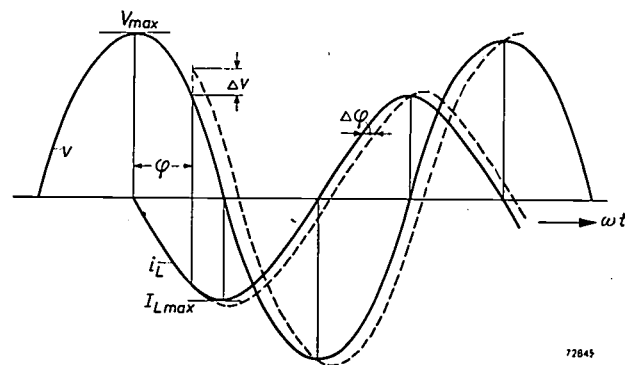


Fig. 2. Diagrammatic representation of the influence of a short interfering pulse upon the momentary amplitude and phase of the oscillator voltage. The phase discontinuity $\Delta \varphi$ is to be particularly noted.

Introducing the relative pulse strength ϵ and the relative amplitude variation x :

$$\epsilon = \frac{\Delta v}{V_{\max}} = \frac{i_s \Delta t}{C V_{\max}}, \dots \dots \dots (7)$$

and

$$x = \frac{\Delta V_{\max}}{V_{\max}} = \frac{\Delta I_{L\max}}{I_{L\max}}, \dots \dots \dots (8)$$

eq. (6) can be written:

$$\left. \begin{aligned} x \cos \varphi - \Delta\varphi \sin \varphi &= \varepsilon, \\ x \sin \varphi + \Delta\varphi \cos \varphi &= 0, \end{aligned} \right\} \dots \dots (9)$$

from which follows:

$$\boxed{\begin{aligned} x &= \varepsilon \cos \varphi, \\ \Delta\varphi &= -\varepsilon \sin \varphi. \end{aligned}} \dots \dots (10)$$

Our further investigation is based upon these equations (10). It is seen that interfering pulses occurring at a voltage peak ($\varphi = 0$ or $\varphi = \pi$) influence the amplitude but not the phase, whereas those occurring when the circuit voltage passes through zero ($\varphi = \pi/2$ or $\varphi = 3\pi/2$) cause only a variation in phase. Further it should be recognized that the amplitude changes are neutralized owing to the limiting mechanism in course of time, but the phase changes remain, or, in other words, it might be said that the oscillator has an "infinite memory" for phase disturbances.

Synchronization by a series of equidistant pulses

The second of the two equations (10) already provides the possibility of synchronization by an external signal, and at the same time it shows the mechanism: when at each cycle of the oscillator voltage an interfering pulse occurs at the phase angle φ , there is a phase discontinuity $\Delta\varphi = -\varepsilon \sin \varphi$ per cycle, so that there is a relative frequency variation $\Delta\varphi/2\pi$. The frequency is increased when the pulses occur prior to the voltage maximum and decreased when they occur after that maximum, as is the case in the position represented in fig. 2.

The resulting relative frequency variation is greatest when the pulses lead or lag by an angle $\pi/2$ with respect to the voltage maximum. Pulses of the relative strength ε can cause at most a relative frequency change $\varepsilon/2\pi$. If the relative difference between the pulse frequency and the undisturbed oscillator frequency is greater than $\varepsilon/2\pi$, no synchronization can take place and a beat phenomenon arises, which will be reverted to later. If the relative frequency difference is less than $\varepsilon/2\pi$, synchronization can take place at a smaller phase angle ($-\pi/2 < \varphi < \pi/2$). As long as the state of synchronism has not yet been reached, the phase φ changes from cycle to cycle until the pulses occur at the right value of φ . Experience teaches that once this is reached the state of synchronism continues, even when there are small variations in the strength or the instant of application of the pulses. It would lead us too far to give here the proof of this stability.

From what has been said it is seen that the synchronization of an LC oscillator by means of pulses

can be represented almost as readily as that of a multivibrator or a blocking oscillator. It could be represented by a synonymous picture of a clock being put back 5 minutes after every 65 minutes, so that it runs synchronously with an interference having a cyclic duration of 65 minutes; as a result the clock shows "hours" of 65 minutes.

Synchronization by a sinusoidal voltage

If the synchronizing signal is not pulse-shaped the phase variation per cycle can be found by integrating over a cycle with the aid of the second equation (10), the signal then being regarded as consisting of elementary pulses. Owing to the factor $\sin \varphi$ in (10), this integration corresponds to a determination of the fundamental component of the disturbing signal. It may therefore be concluded that the synchronizing action of a periodically repeated pulse is equal to that of a sinusoidal signal of the same strength as the fundamental component of the series of pulses. The synchronization by a sinusoidal signal can thus be derived in a simple way from that by pulses.

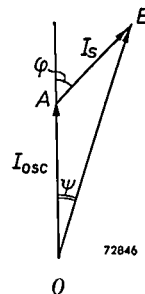


Fig. 3. The vector OA, which is in phase with the circuit voltage, represents the anode alternating current of the oscillator valve. The vector AB represents the synchronizing sinusoidal interfering current. The angle φ is the phase angle between the latter and the circuit voltage, whilst the angle ψ is a measure for the frequency deviation of the oscillator.

The case of synchronization by a sinusoidal signal can also be deduced from the vector diagram of fig. 3. The alternating anode current i_{osc} of the oscillator valve is in phase with the circuit voltage. The signal current i_s is not in phase with it. The vector OB represents the total current fed to the LC circuit. The constant phase angle φ between the latter and the circuit voltage is only possible if the oscillator frequency differs from the resonant frequency of the circuit. The diagram does not show how the synchronization is achieved; it only represents the possible final state⁵⁾.

Differential equation for the total phase deviation under the influence of a sinusoidal disturbance

In order to investigate the beat phenomenon arising when the interfering signal is too weak to bring about synchronization, a differential equation will be derived for the general case of a sinusoidal interference whose angular frequency differs by an amount $\Delta\omega$ from that of the undisturbed oscillator.

⁵⁾ An article on synchronization and the beat phenomenon, in which this diagram is dealt with, was given by R. Adler, Proc. Inst. Rad. Engrs. 34, 351-357, 1946.

The phase deviation of the oscillator voltage under the influence of the interference is called Φ , which must be considered as a function of time.

The oscillator voltage is therefore assumed to be

$$v = V_{\max} \cos(\omega_0 t + \Phi), \dots (11)$$

and the interfering current fed to the LC circuit is put:

$$i_s = I_{s \max} \cos(\omega_0 t - \Delta\omega t - \varphi) \dots (12)$$

The term $\Delta\omega$ is small compared with ω_0 , whilst φ is a constant phase angle as yet undetermined. The momentary phase difference between the interference and the oscillator voltage is clearly

$$\Theta = \Delta\omega t + \varphi + \Phi. \dots (13)$$

The interfering current is split up into a component which is in phase with the oscillator voltage and one which shows a phase difference of $\pm\pi/2$. The first component influences the amplitude but not the phase of the oscillator, while the reverse applies for the second component. By employing a known trigonometric relation we find from (12) that:

$$i_s = I_{s \max} \cos(\omega_0 t + \Phi) \cos(\Delta\omega t + \varphi + \Phi) + I_{s \max} \sin(\omega_0 t + \Phi) \sin(\Delta\omega t + \varphi + \Phi). (14)$$

Here the last product represents the interfering component which is in quadrature with the momentary oscillator voltage and thus responsible for the phase variations. It is only this component that will now be considered.

In order to arrive at the differential equation referred to, the momentary interfering current has to be multiplied, in accordance with the second equation (10), by the sine of the phase angle through which the oscillator voltage has passed subsequent to the last voltage maximum. From (11) it appears that, to within a multiple of 2π , this angle is equal to $\omega_0 t + \Phi$. Hence, from (7), (10) and (14):

$$\begin{aligned} d\Phi &= -\frac{I_{s \max} dt}{CV_{\max}} \sin^2(\omega_0 t + \Phi) \sin(\Delta\omega t + \varphi + \Phi) = \\ &= \frac{I_{s \max} dt}{2CV_{\max}} [1 - \cos 2(\omega_0 t + \Phi)] \sin(\Delta\omega t + \varphi + \Phi). \end{aligned}$$

From this equation it is seen that the phase deviation Φ is subject to a gradual variation relating to the frequency difference $\Delta\omega$, but that Φ also shows a rapid variation within each cycle of the oscillator voltage. This rapid variation thus bears more the character of a variation of the wave form of the oscillator voltage, rather than that of a real phase fluctuation. The corresponding effect can therefore be ignored, so that finally we arrive

at the following differential equation:

$$\frac{d\Phi}{dt} = -\frac{I_{s \max}}{2CV_{\max}} \sin(\Delta\omega t + \varphi + \Phi). \dots (15)$$

A closer consideration reveals that the rapid phase variation within each cycle has indeed no physical significance but is only related to the fact that we have taken as the undisturbed phase that of a freely oscillating circuit, and not that of a circuit driven by a sinusoidal current with the frequency ω_0 . Thus the term ignored was only a complication due to the simplified treatment of the problem.

The case of synchronization is easily deduced from the differential equation (15), for then it is obvious that

$$\Phi = -\Delta\omega t.$$

Substituting this in (15) we find:

$$\frac{I_{s \max} \sin \varphi}{2CV_{\max} \Delta\omega} = 1, \dots (16)$$

where, in accordance with fig. 2 and fig. 3, φ represents the phase angle between the interfering current and the oscillator voltage. The condition for synchronization is clearly:

$$\frac{I_{s \max}}{2CV_{\max} \Delta\omega} \geq 1. \dots (17)$$

The beat phenomenon in the case of a very weak interference; comparison with a normal beat

When the interference is very weak the phase of the oscillator will deviate very little from that in the case of no interference. An approximate solution of (15) is then obtained by ignoring the term Φ on the right-hand side, thus giving as a first approximation for a weak interference:

$$\Phi = \frac{I_{s \max}}{2CV_{\max} \Delta\omega} \cos(\Delta\omega t + \varphi). \dots (18)$$

This equation represents a phase modulation with the angular frequency $\Delta\omega$, the difference between the angular frequency of the signal and that of the oscillator. When compared with (17) it shows that the maximum phase deviation must remain below 1 radian, as otherwise (18) is certainly no longer valid as a solution of (15). The exact solution of (15) will be reverted to later.

It is very instructive to compare the beat phenomenon represented by (18) with a normal beat between two fixed frequencies. This is done by replacing the oscillator current i_{osc} in fig. 1 by an equally strong current with the constant frequency ω_0 . This gives across the LC circuit a voltage of the amplitude

$$V_{\max} = I_{osc \max} R, \dots (19)$$

where R is the parallel resistance of the circuit at resonance. With the series resistance r this bears the known relation:

$$R = \frac{L}{Cr} \dots \dots \dots (20)$$

The interfering current represented by (12) gives, across the LC circuit, the voltage amplitude

$$V_{s \max} = I_{s \max} \frac{R}{\sqrt{1 + y^2}}, \dots \dots (21)$$

where y has the value

$$y = \frac{2L}{r} \Delta\omega = 2RC \Delta\omega \dots \dots (22)$$

The sum of the voltage vectors given by (19) and (21) can be found in the known way from a diagram in which $V_{s \max}$ has the relative angular velocity $\Delta\omega$ with respect to V_{\max} (fig. 4).

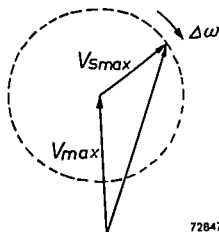


Fig. 4. Vector diagram of a normal beat. With respect to V_{\max} the vector $V_{s \max}$ has the relative angular velocity $\Delta\omega$.

If $V_{s \max}$ is sufficiently small compared with V_{\max} it is clear that, compared with the case where there is no interference, for the normal beat the maximum phase deviation is

$$\Phi_{\text{normal}} = \frac{V_{s \max}}{V_{\max}} \dots \dots (23)$$

Finally, from (18) to (23) inclusive, we find for the proportion of the maximum phase deviations for the oscillator and for the normal beat:

$$\frac{\Phi_{\text{osc}}}{\Phi_{\text{normal}}} = \frac{\sqrt{1 + y^2}}{2CR\Delta\omega} = \frac{\sqrt{1 + y^2}}{y} \dots (24)$$

This shows that, particularly for small values of $\Delta\omega$ ($y < 1$), the phase deviation of an oscillator is much greater than that of a normal beat. This phenomenon can be further explained in the following way.

When a disturbance of short duration is applied to an LC circuit fed by a current of constant frequency, a change takes place in the momentary phase. After a time this change decays owing to the

circuit damping. In the case of an oscillator, however, this phase disturbance continues to exist for an indefinite time. Consequently (below the limit set by the occurrence of synchronization) in the case of a beat whose cycle is longer than the decay time $2L/r$ of the circuit the phase deviation of an oscillator may become much larger than in the case of a normal beat. This is undoubtedly a remarkable and little known property of an oscillator.

Exact solution of the differential equation (15)

Although a fairly good general picture has already been formed, it has not yet been made clear how in the case of an interference growing in strength the beat phenomenon is transformed into the state of synchronization. The answer to this question lies in the exact solution of (15). It will appear that the beat phenomenon itself is already accompanied by an initially very small change in the mean oscillator frequency.

At first sight solving the differential equation (15) appears to be rather complicated, but it can nevertheless be accomplished with elementary means. For this purpose we introduce as a new variable:

$$u = \Delta\omega t + \varphi + \Phi, \dots \dots (25)$$

and for the sake of brevity we furthermore put:

$$\frac{I_{s \max}}{2CV_{\max}\Delta\omega} = a \quad (a < 1), \dots \dots (26)$$

so that (15) becomes:

$$\frac{du}{1 - a \sin u} = \Delta\omega dt.$$

The left-hand member of this equation leads to, an elementary trigonometric integral, by means of which we find as the solution of (15):

$$\frac{2}{\sqrt{1 - a^2}} \arctan \left[\sqrt{\frac{1+a}{1-a}} \tan \left(\frac{u}{2} - \frac{\pi}{4} \right) \right] = \Delta\omega t + \text{const.}$$

By putting $\Phi = 0$ at a suitably chosen moment, and by employing eq. (25), this equation can, without loss of generality, be written:

$$\tan \left(\frac{\Delta\omega t}{2} + \frac{\Phi}{2} \right) = \sqrt{\frac{1-a}{1+a}} \tan \left(\frac{\sqrt{1-a^2}}{2} \Delta\omega t \right). \quad (27)$$

This in turn can be abbreviated to read:

$$\tan \psi_2 = \sqrt{\frac{1-a}{1+a}} \tan \psi_1, \dots \dots (28)$$

where ψ_1 and ψ_2 stand for the respective terms in brackets in (27).

From (28) it follows that $\tan \psi_1$ and $\tan \psi_2$ assume simultaneously the value 0 and likewise simultaneously the value ∞ . Thus the angles ψ_1 and ψ_2 reach the values 0, $\pi/2$, π , $3\pi/2$ at the same time. This is graphically represented in fig. 5. On an average the angles ψ_1 and ψ_2 obviously increase at the same rate. Therefore, disregarding the periodic fluctuation, the average increase of Φ is given by:

$$\Phi = -\Delta\omega t (1 - \sqrt{1 - a^2}).$$

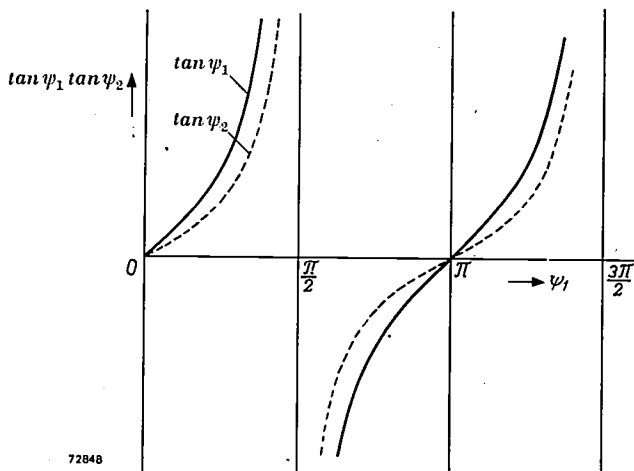


Fig. 5. Graphic representation of the equations (28) and (27). The angles ψ_1 and ψ_2 increase, on an average, at the same rate.

and the average angular frequency difference between the interference and the oscillator is obviously:

$$\Delta\omega \sqrt{1 - a^2}.$$

This is in agreement with the known phenomenon that for a not too weak interference the beat note is perceptibly lower than the difference $\Delta\omega/2\pi$ between the frequencies of the interference and the undisturbed oscillator. Thus the mean oscillator frequency is "pulled" by the interference before there is a possibility of synchronization.

From (28) and fig. 5 it can moreover be deduced that, when a approaches unity, the maximum difference between ψ_2 and ψ_1 approaches $\pi/2$. Since the fluctuation in the increase of ψ_2 corresponds to that of $\Phi/2$, it is seen that, as the possibility of synchronization is approached ($a \rightarrow 1$), the phase fluctuation of the oscillator due to the decreasing beat frequency becomes very slow and approaches the maximum amplitude $\pm\pi$. Thus the oscillator voltage passes through each phase relation with respect to the interference, so that as a increases further, the state of synchronization can come about without any discontinuity. This completes the picture of the phenomena occurring.

Non-linearity of the phenomena; application to arbitrary disturbances; the frequency noise of oscillators

Considering that in the foregoing the non-linearity of the amplitude-limiting mechanism does not play any particular part, it might be thought that a linear treatment of the interference problem has been given. This, however, is by no means the case, as already follows from the unmistakable non-linearity of the differential equation (15).

It is true that the approximative solution (18) is linear, so that it can be applied to a sum of two or more sinusoidal interferences, provided that for each interference the value of a defined by (26) is sufficiently small with respect to unity. If this condition is not satisfied, then the effect of an interference is no longer independent of the presence of other interferences. This is most evident for the case of synchronization: synchronization by two periodic interferences of different frequencies is a *contradictio in terminis*.

In the case of some non-periodic disturbance, for instance a source of noise in the oscillator itself, it might be suggested to take the Fourier spectrum of that disturbance and to apply to its components the linear approximate solution (18). This, however, is not permissible, since as a rule the spectrum will contain components for which the condition $a \ll 1$ is not satisfied. And these happen to be the most interesting components, namely those which are grouped immediately around the undisturbed oscillator frequency⁶⁾.

Another method, and in our opinion this is the only correct one, is to start from the time function of the disturbance, thus in the case of a noise source from the real statistical fluctuations. One then arrives at a mean value of the phase fluctuation per cycle. This will not be gone into further here because we hope to revert to this subject elsewhere⁷⁾. In the case of arbitrary disturbances which are given functions of time, as a rule the differential equation corresponding to (15) will have to be solved, which in general will not be possible by elementary means.

⁶⁾ If an oscillator is to be "linearized" by giving the controlling mechanism a very large time constant, it is not permissible to introduce at the same time a constant oscillator voltage.
⁷⁾ A treatise on oscillator noise, based upon the statistical fluctuations, has meanwhile been given by Aug. Blaquièrre, Comptes Rendus Ac. Sc. Paris 234, 419-421, 710-712, 1140-1142, 1952 (Nos. 4, 7 and 11). Articles based on the Fourier spectrum of the noise source have been published by A. Spaelti, Bull. Schweiz. Elektrot. Vereins 39, 419-428, 1948 and L. M. Lerner, Proc. Nat. Electronics Conference 7, 275-287, 1951.

Summary. Starting from the supposition that the voltage produced by an oscillator is approximately sinusoidal and that the amplitude-controlling mechanism has no effect upon the phase of the oscillator voltage, the phase shift of that voltage due to an interfering pulse has been calculated. This is followed by an elementary explanation of the synchronization of an oscillator by a periodic disturbance differing in the duration of its cycle from that of the oscillator. By solving a differential equation the phase and frequency variations under the in-

fluence of a weak sinusoidal interference are calculated. As the strength of the interference increases, or the frequency difference decreases, the solution passes continuously into the state of synchronization. Further, the non-linearity of the phenomena is discussed, i.e. the dependency of the effect of an interference upon the presence of other interferences. This is followed by an explanation how the method followed can be applied for calculating the influence of noise voltages in the oscillator upon the frequency constancy of the oscillator.

PHOTOSYNTHESIS

by R. van der Veen.

581.132.1

In the final analysis, all life on this earth depends on the transformation of sunlight into chemical energy. This process, called photosynthesis, takes place exclusively in green plants and even there the output is small. If in this respect mankind could succeed in making itself independent of the vegetable world, the prospects for the energy provision of the whole world would be very favourably influenced. If we really intend to reach this goal it is important to obtain a fuller insight into the complicated process of photosynthesis; at present only the fundamental principles are known. It has been found that we can contribute to our knowledge by studying the fluorescence of living elements of the vegetable world.

The energy radiated from the sun to the earth amounts to about 10^{18} kWh a year. The energy consumption in a country as highly developed as the USA amounts only to about 10^{13} kWh a year, while the figure for the whole world is only 3 times as large as that for the USA. By far the largest part of the energy used by mankind has been derived since time immemorial from the coal and oil deposits found in the earth's crust, formed long ago by solar energy. These deposits however are not inexhaustible, and assuming that present-day technical development proceeds at the same speed, it has been calculated that after some centuries these deposits will have been completely consumed. (If we tried to replace the solar energy supplied to us every day by the energy stored in the oil and coal deposits we would consume the whole stock within 3 days!)

As might be expected science is already carrying out experiments to compensate the effects of this depletion. A modest step in this direction has been made by using atomic energy released from thorium and uranium but the available quantities of these elements are limited. If we could succeed in transmuting hydrogen into helium in a technically workable process, we certainly would postpone the day of exhaustion for many hundreds of thousands of

years, but to achieve this we have still a long way to go.

It is obviously desirable to exploit the possibility of making direct use of solar energy. A small percentage of the energy in the world is generated in this way, viz. by making use of wind and water power, but it is improbable that the amount of energy so obtained can be considerably increased. Now, it is known that part of the total solar energy supplied is transformed by green plants into chemical energy. This process — photosynthesis — is the basis of all life and consists of the formation of carbohydrates from water and the carbonic acid of the atmosphere by the influence of sunlight. This photosynthesis results in the accumulation of energy to an amount of 10^{15} to 10^{16} kWh a year. However, this energy is lost again for the greater part without being used to an appreciable extent by mankind, by decomposition of the organic substances through bacteria, etc.

If we look at these figures it is evident that all our worries about the exhaustion of our energy deposits would disappear at once if we could only succeed in making a more efficient use of this photosynthesis. The first step in this direction is to obtain an insight into the very complicated system of reactions which constitute photosynthesis.

Summary. Starting from the supposition that the voltage produced by an oscillator is approximately sinusoidal and that the amplitude-controlling mechanism has no effect upon the phase of the oscillator voltage, the phase shift of that voltage due to an interfering pulse has been calculated. This is followed by an elementary explanation of the synchronization of an oscillator by a periodic disturbance differing in the duration of its cycle from that of the oscillator. By solving a differential equation the phase and frequency variations under the in-

fluence of a weak sinusoidal interference are calculated. As the strength of the interference increases, or the frequency difference decreases, the solution passes continuously into the state of synchronization. Further, the non-linearity of the phenomena is discussed, i.e. the dependency of the effect of an interference upon the presence of other interferences. This is followed by an explanation how the method followed can be applied for calculating the influence of noise voltages in the oscillator upon the frequency constancy of the oscillator.

PHOTOSYNTHESIS

by R. van der Veen.

581.132.1

In the final analysis, all life on this earth depends on the transformation of sunlight into chemical energy. This process, called photosynthesis, takes place exclusively in green plants and even there the output is small. If in this respect mankind could succeed in making itself independent of the vegetable world, the prospects for the energy provision of the whole world would be very favourably influenced. If we really intend to reach this goal it is important to obtain a fuller insight into the complicated process of photosynthesis; at present only the fundamental principles are known. It has been found that we can contribute to our knowledge by studying the fluorescence of living elements of the vegetable world.

The energy radiated from the sun to the earth amounts to about 10^{18} kWh a year. The energy consumption in a country as highly developed as the USA amounts only to about 10^{13} kWh a year, while the figure for the whole world is only 3 times as large as that for the USA. By far the largest part of the energy used by mankind has been derived since time immemorial from the coal and oil deposits found in the earth's crust, formed long ago by solar energy. These deposits however are not inexhaustible, and assuming that present-day technical development proceeds at the same speed, it has been calculated that after some centuries these deposits will have been completely consumed. (If we tried to replace the solar energy supplied to us every day by the energy stored in the oil and coal deposits we would consume the whole stock within 3 days!)

As might be expected science is already carrying out experiments to compensate the effects of this depletion. A modest step in this direction has been made by using atomic energy released from thorium and uranium but the available quantities of these elements are limited. If we could succeed in transmuting hydrogen into helium in a technically workable process, we certainly would postpone the day of exhaustion for many hundreds of thousands of

years, but to achieve this we have still a long way to go.

It is obviously desirable to exploit the possibility of making direct use of solar energy. A small percentage of the energy in the world is generated in this way, viz. by making use of wind and water power, but it is improbable that the amount of energy so obtained can be considerably increased. Now, it is known that part of the total solar energy supplied is transformed by green plants into chemical energy. This process — photosynthesis — is the basis of all life and consists of the formation of carbohydrates from water and the carbonic acid of the atmosphere by the influence of sunlight. This photosynthesis results in the accumulation of energy to an amount of 10^{15} to 10^{16} kWh a year. However, this energy is lost again for the greater part without being used to an appreciable extent by mankind, by decomposition of the organic substances through bacteria, etc.

If we look at these figures it is evident that all our worries about the exhaustion of our energy deposits would disappear at once if we could only succeed in making a more efficient use of this photosynthesis. The first step in this direction is to obtain an insight into the very complicated system of reactions which constitute photosynthesis.

Photosynthesis

Photosynthesis is the most important chemical reaction taking place on earth. All life is dependent on the chemical energy stored in its own organic substance. The processes which govern life are kept going by the release of this energy by respiration, i.e. by the combustion of organic substance. Consequently a primary condition for the conservation of life is the formation of organic matter in one way or another. This formation takes place mainly by a single method, viz., by the influence of sunlight on a certain substance present in green plants — chlorophyll. (Certain organisms capable of using light for building organic substances are not green but close examination has shown that they do contain chlorophyll, the green colour being suppressed by other colours.)

It is remarkable that the higher developed beings such as the animals, are unable to make direct use of photosynthesis. Not even in the laboratory has this proved possible. This means that all the higher developed creatures including mankind have to live parasitically on the green plants.

In general the chlorophyll present in the leaves of a plant is in the form of granules or chloroplasts, moving about in the plasma of the cell. In *fig. 1* an electron micrograph of some chloroplasts is shown. It is clearly to be seen that each

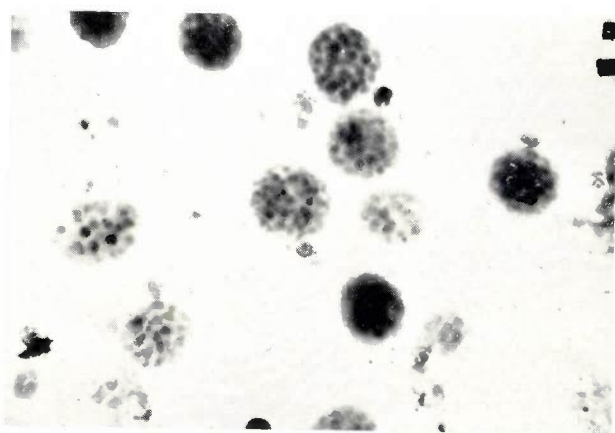


Fig. 1. Electron micrograph of chlorophyll granules (chloroplasts). The distance between the black squares in the right-hand top corner = 1μ . It can be seen that the granules are built up from small corpusculae, the grana, embedded in the colourless stroma.

chloroplast consists of a number of lenticular bodies, the grana, embedded in a colourless gel, the stroma. *Fig. 2* shows a fragment of a chloroplast at a higher magnification.

The photosynthesis process as such takes place inside the chloroplasts. Schematically, photosyn-

thesis can be represented by the reaction:



Though this formula looks very simple, the process that really happens is extremely complicated and



Fig. 2. Electron micrograph of a fragment of a chlorophyll granule. Four grana are visible, embedded in a patch of stroma. The distance between the white squares in the right-hand bottom corner = 1μ .

up till now it has been only partly unravelled. The final product, represented by (CH_2O) , and which should be read as a collection of higher carbohydrates (sugars) is the organic substance which supplies the energy for the conservation of life.

Thus in fact photosynthesis consists of the reduction of CO_2 to carbohydrates, whereas the respiration process corresponds to the inverse reaction, i.e. the oxidation of carbohydrates to CO_2 .

Mechanism of the reaction

It has been found that the chloroplasts alone cannot effect a photosynthesis of carbohydrates from CO_2 and water. After having been exposed to light, the chloroplasts do possess a certain reducing power, as can be concluded from the reduction of easily reducible substances as e.g. ferric salts to ferrous salts, during which process oxygen is liberated. In order to reduce carbon dioxide, however, a preliminary treatment of this substance is necessary. This preliminary treatment takes place in the protoplasm, independent of the light; for this reason it is called a dark-reaction. The light reaction proper, coupled with a number of dark-reactions, takes place in the chloroplasts.

From the foregoing it is already evident that the photosynthesis process as a whole is infinitely more complicated than is suggested by the simple equat-

ion given above. To clarify the picture — without undue complication — *fig. 3* illustrates how the reaction system looks according to present-day knowledge.

In many (largely unknown) intermediate stages the CO_2 from the atmosphere is converted into a

have succeeded in gaining a better insight into the various reactions which take place.

The first author has shown that under special circumstances it is possible to effect the CO_2 conversion by the chloroplasts, by adding diphosphopyridinenucleotide (DPN) to a suspension

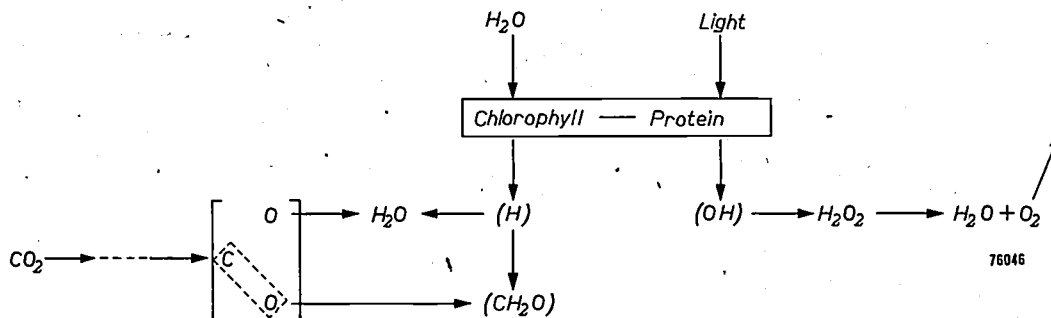


Fig. 3. Schematic reproduction of the photosynthesis process. The dark-reactions are on the left. On the right is the reduction of H_2O_2 to H_2O and O_2 by the enzyme catalase. The actual light reaction in the chloroplast narrows down to the decomposition of water into an (H)-component and an (OH)-component. All the oxygen which finally escapes originates from the water, not from absorbed CO_2 .

complexity of unknown chemical compounds, indicated in the scheme of *fig. 3* as $[\text{C}_0^{\text{O}}]$. At the same time water is decomposed by the chlorophyll-protein complex of the chloroplasts under the influence of light into a reducing component, indicated by the symbol (H) and an oxidizing component indicated by (OH). It has not yet been agreed as to what is meant exactly by these compounds; it is possible that we have to deal with H-atoms and OH-radicals.

The (H)-compound serves to reduce the $[\text{C}_0^{\text{O}}]$ complex, formed in the dark-reaction, to carbohydrates and water. The (OH)-components probably combine to H_2O_2 -molecules, which, under the influence of the enzyme catalase (a substance, which has been isolated and is present in a relatively high concentration in every chloroplast) decompose immediately in water and oxygen. The oxygen escapes.

According to this scheme, the oxygen which finally escapes originates entirely from the water in the plant. From the symbolic equation at the beginning of this paper one might expect that half the liberated oxygen — or even more — originates from the CO_2 , but experiments carried out with the oxygen isotope O^{18} have completely disproved this simple conception.

Much research work has been carried out to clarify those details of the photosynthesis which are still obscure. Among others, Ochoa¹⁾ and Calvin²⁾

of chloroplasts. DPN is a complicated, phosphorus-containing organic substance, which seems to be present in all live cells and which is reducible. However, when other reducible substances and certain enzymes are present, it is easy to re-oxidize the reduced DPN. Ochoa showed that in the presence of reduced DPN and certain enzymes, CO_2 can combine with pyruvic acid to form malic acid. During this reaction the reduced DPN is oxidized, but it is immediately reduced again by the chloroplast which has been exposed to light. Thus, in a sense, the DPN acts as a catalyst. By the combined action of chloroplast, light energy, DPN and enzymes, the reaction always proceeds in the same direction, i.e. in the direction of the reduction of CO_2 to carbohydrates.

By making use of radio-active carbon, Calvin has proved, however, that the CO_2 fixed during the photosynthesis is first found, not in malic acid, but in the form of phosphoglyceric acid. After some time the radio-active C-atoms spread over a large number of other compounds. It may well be that these results are not opposed to those of Ochoa: in the sense that both results are based on the reduction of DPN by the chloroplasts. It is interesting to note that organic phosphates are necessary to effect photosynthesis, a fact which has been established by Wassink, Winternans and Tjia³⁾.

¹⁾ S. Ochoa, *Symposia Soc. Exp. Biol.* 5, 29, 1951.

²⁾ M. Calvin, *J. Chem. Educ.* 26, 639, 1949.

³⁾ E. C. Wassink, J. F. G. M. Winternans and J. E. Tjia, *Proc. Kon. Akad. Wet. Amst., Series C*, 54, 496-502, 1951 (No. 5).

The question arises — why is the course of the photosynthesis reaction so extremely complicated? To answer this question it should be realized that in nature only those processes which lead to the final product in an efficient manner have a chance to succeed. In order to effect so radical a reaction as the synthesis of carbohydrates by means of the relatively small energy (light and heat) available, nature has to rely on subdividing the reaction into a very large number of intermediate reactions, the direction (sense) of which can be easily influenced by the circumstances. Now the presence of light and certain enzymes results in a slight shift of the equilibria in the plant cells towards the reducing direction. This requires only a very small energy-absorption per process. Moreover, it should be taken into account that the compounds formed not only play a part in the reactions considered, but also have very definite functions in the organisms, which results in the increase of the "efficiency" of the entire process.

Light saturation

The speed of photosynthesis is determined by the slowest reaction in the chain of the light and dark-reactions. The rate of progress of the light-reaction increases with the increase of the light intensity. In the temperature region where physiological processes can normally take place, i.e. between about 5—30 °C, this speed is almost independent of the temperature. The dark-reactions, on the contrary, are pure chemical reactions, of which the

rate of progress rapidly increases with the temperature. At a given temperature an increase of the light intensity will result in an increase of photosynthesis, but only as long as the light-reaction is the limiting factor. When one of the dark-reactions becomes limiting, an increase of the light intensity can no longer cause a change in the rate of photosynthesis: light saturation sets in. The higher the temperature, the greater the rate of progress of the dark-reactions, and light saturation will set in at higher intensities. This has been represented schematically in *fig. 4*.

Analysis of the light-reaction by means of fluorescence

As already demonstrated, only the fundamental principles of the process of photosynthesis are understood at present. The transitions which take place in the dark-reactions are still unknown and so is the light-reaction. But there is a possibility of becoming better acquainted with the latter.

Only part of the absorbed light energy is transformed in chemical energy during the light-reaction. In nature the output is only a few percent, whereas it has been possible to obtain an output of about 30% or more in the laboratory when working with low light intensities. The remainder of the light energy is transformed into heat and for a small part into fluorescent radiation. The latter fact is most fortunate, for by studying the fluorescence, it is possible to follow the course of the light-reaction to a certain extent. It may be assumed that the relation between the energy given off as heat and the energy given off as fluorescence is always constant, whatever the percentage of the absorbed radiation energy used for the chemical transformation. Consequently the intensity of the fluorescence is a yardstick for the quantity of light energy bound chemically, and so for the output of the photosynthesis: the intensity of the fluorescence will be smaller according as the output is higher.

It is evident that the measurement of the intensity of the fluorescence can teach us only about the progress of the light-reaction. As far as the progress of the dark-reactions (including those following the light reaction) are concerned, this method will bring us no further. It may well be that part of the chemically bound energy is liberated in a later stage of the process as heat. This is the case e.g. if a leaf is exposed to light in the absence of CO₂. In that case there is no material that can be reduced and no oxygen is set free. However, the fluorescence is hardly any higher than it would be if CO₂ were

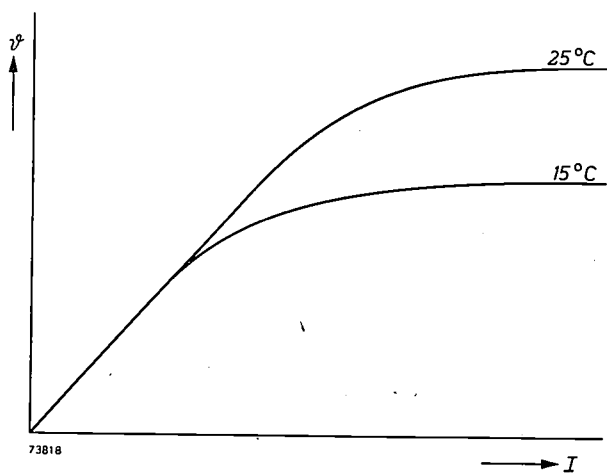


Fig. 4. Speed v of photosynthesis as a function of the illumination intensity I and the temperature. The light-reaction determines the speed in the sloping part of the curve, whereas the speed of the reaction is completely limited by the dark-reactions in the horizontal part, where light saturation sets in. The rate of progress of these dark-reactions increases as the temperature becomes higher. The light-reaction is almost independent of the temperature, the result being that light saturation at a higher temperature sets in only at a higher light intensity.

present. The fundamental light-reaction, i.e. the splitting-up of water into (H)- and (OH)-compounds, obviously takes place at the same rate, the presence or absence of CO_2 being immaterial. However, CO_2 being absent, these compounds combine again to water, during which process heat is generated and from a chemical point of view, it appears that nothing has happened.

The Philips Laboratory has carried out extensive researches on the phenomena which occur at the beginning of the illumination in the process of photosynthesis. In particular the CO_2 -absorption and the fluorescence were studied⁴⁾.

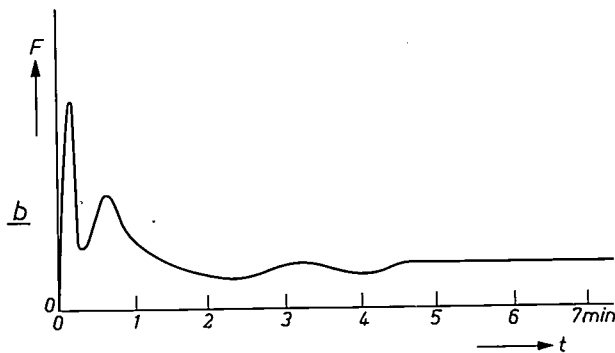
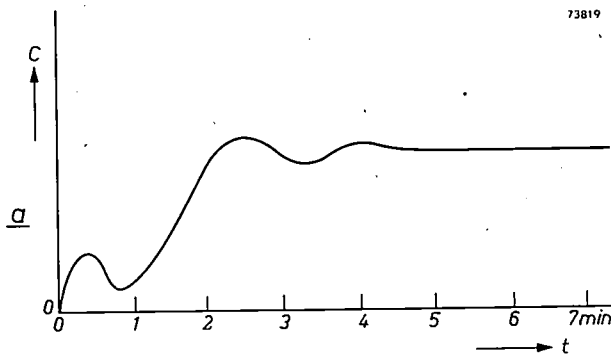


Fig. 5. a) Progress of the CO_2 -absorption C per unit of time. b) Intensity of the fluorescence F , as function of the time t after applying the illumination. With the exception of the very early part, the two curves are nearly reflected images.

The general course of the CO_2 -absorption is illustrated in fig. 5a. During the first moments (about 10 seconds) rather a large quantity of CO_2 is absorbed, but then the rate of absorption decreases quickly. About a minute after illumination is started, there is a gradual increase of the photosynthesis (adaptation period), until an almost constant level of the CO_2 -absorption is reached, the height of the level being dependent on the intensity of illumination and the plant in question. The fluorescence shows, apart from the early beginning period, closely corresponding variations

⁴⁾ See, for example, R. van der Veen, *Physiologia Plantarum* 4, 486-494 1951 (No. 3).

(fig. 5b): the CO_2 absorption at the beginning is reflected in a short-period decrease, and the adaptation period as a continued gradual decrease. Fluctuations in the photosynthesis which may set

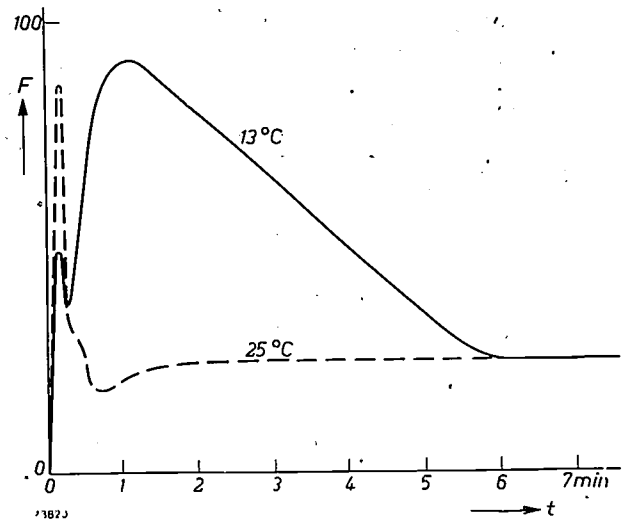


Fig. 6. Intensity of fluorescence F as function of the time t after applying the illumination, at two temperatures, 13°C and 25°C . Though the ultimate level is the same in both cases, it is attained much earlier at the higher temperature. The measurements have been carried out on a dahlia-leaf.

in later are completely reproduced in the intensity of the fluorescence. In the very early part, however, the fluorescence shows a sharp peak, the relation of which to the photosynthesis is still obscure.

The influence of the temperature, the light intensity and the duration of the dark period preceding the illumination has been investigated. In the early stages the rate of photosynthesis is much lower at low temperatures (say, 13°C) than at high

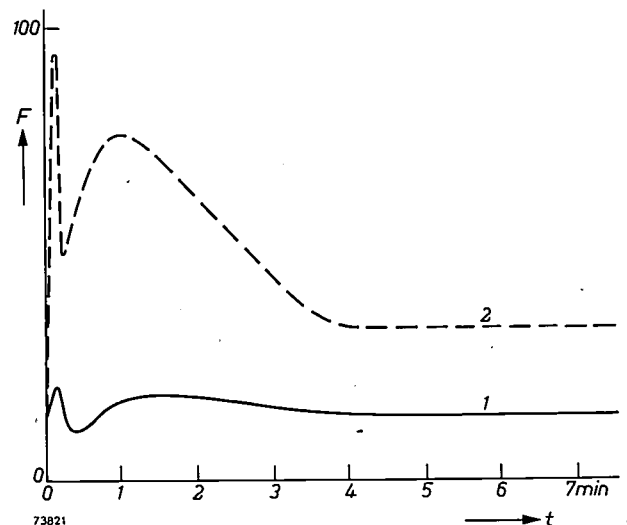


Fig. 7. Intensity of fluorescence F as function of the time t after applying the illumination at two intensities, differing by a factor 3. Curve 1 appertains to the low intensity, curve 2 to the high one. Measurements have been carried out on a tomato leaf.

temperatures (25 °C). This is reflected in the high value of the fluorescence at 13 °C as compared with that at 25 °C, during the first minutes (*fig. 6*). It can be seen, however, that the ultimate level of the fluorescence is not influenced by the temperature. This confirms the assertion that the light reaction depends little on the temperature.

The intensity of the fluorescence increases when the light intensity is increased, as would be expected. It is remarkable, however, that the increase is much more marked when the illumination is first begun than in later periods (*fig. 7*).

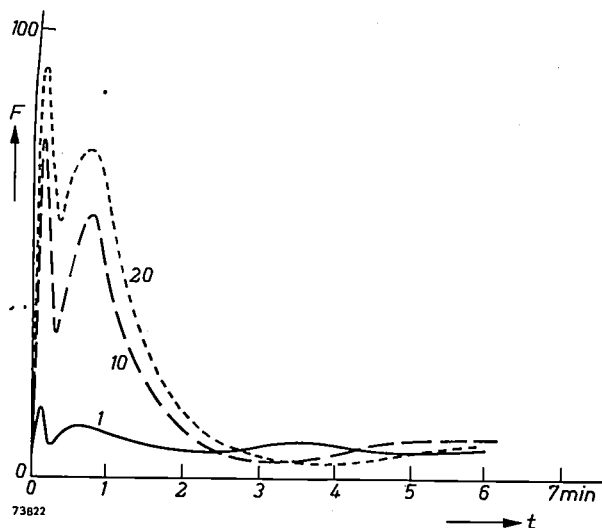


Fig. 8. The influence of the length of the preceding dark period on the fluorescence curve. The figures against each curve represent the length of the period in minutes. The measurements were carried out on a tomato leaf.

The influence of the preceding dark period is represented in *fig. 8*. The longer the period, the higher are the various maxima. This agrees completely with what has been found by other, more direct means. It seems that the plant remembers the previous illumination, provided this took place not longer than 20 minutes previously.

It is impossible as yet to have a clear picture of the physical and chemical background of all these phenomena, but it is certain that the study of fluorescence is of material aid to investigations into photosynthesis.

Summary. Photosynthesis, the basis of all life on earth, takes place only in green plant parts. The substance called chlorophyll is an essential link in the transformation process of light into chemical energy. The process of photosynthesis, essentially consisting of the reduction of CO_2 to carbohydrates and accompanied by the liberation of oxygen — and as such to be considered as the inverse of the respiration process — is very complicated. The actual light-reaction takes place in the chloroplasts of the plant, but it is preceded by a large series of dark-reactions, occurring partly in the protoplasm, which effect a preliminary transition of the CO_2 into an easily reducible substance. The light-reaction can be considered as the splitting up of water, under the influence of light, into a reducing and an oxidizing compound; the former effecting the reduction of CO_2 to carbohydrates and water, and the latter ultimately liberating oxygen. It has been found that the presence of organic phosphates in the dark-reactions is very important. The output of the photosynthesis in nature is only a small percentage, whereas in the laboratory, at low light intensities, an output of 30% and higher has been attained. A large part of the absorbed light energy is lost as heat, and a small part as fluorescent radiation. By studying the fluorescence of the green parts of plants, science has succeeded in investigating the early part of the light-reaction. The results of the fluorescence research indicate that the intensity of fluorescence is often a measure of the process of photosynthesis.

AN OSCILLATOR WITH CONSTANT OUTPUT VOLTAGE

by L. ENSING and H. J. J. van EYNDHOVEN.

621.396.615-621.316.722.1

Using the general principle of a regulating system, a valve oscillator can be made to work in such a way that the output voltage remains substantially constant. An application regarding an installation for the calibration of valve voltmeters is described.

For the generation of alternating voltages with fixed or controllable frequency, use is often made of oscillators based on an electronic valve (triode or pentode), an oscillatory circuit (LC circuit) and a feedback element. An oscillation generated in this way can, however, show fluctuations in amplitude and in frequency and deviations from the sinusoidal waveform which are most undesirable, especially for measuring and calibrating purposes.

We shall describe below a method which is primarily directed towards making the amplitude of the oscillator independent of disturbing influences such as changes in the supply voltages of the valve, in the mutual conductance of the valve and in the impedance of the LC circuit. With this method the valve is working under conditions which are also favourable for the frequency to remain constant and for the generation of a purely sinusoidal voltage.

Regulating principle applied to an oscillator

A method on which many regulating systems are based¹⁾ consists in the comparison of the regulated quantity x with a similar but constant reference quantity a_{ref} , in that the difference $a_{\text{ref}} - x$ is made to control the apparatus in such a way that $|a_{\text{ref}} - x|$ is reduced.

If the quantity which has to be kept constant and the reference quantity are not similar, then an auxiliary quantity will have to be derived from the first mentioned one which is similar to a_{ref} . This is the case in the matter under discussion: the alternating output voltage of a valve oscillator has to be kept constant in amplitude (V_0), but as reference voltage we can of course only take a direct voltage (E_{ref}). By rectification a direct voltage E_0 which is proportional to V_0 can be derived from the oscillator voltage and the difference $E_{\text{ref}} - E_0$ can function as control voltage, i.e. it can control the oscillator in such a way that the difference $|E_{\text{ref}} - E_0|$, irrespective of its cause, becomes smaller. To attain this, the control voltage can work, for example, as a bias in the control-grid circuit with such polarity

that the working point on the valve characteristic is shifted to a region of greater or smaller mutual conductance accordingly as the voltage amplitude V_0 was either too small or too large.

Influence of the valve adjustment

Before elaborating on this thought it should be recognized that the limitation of the oscillation amplitude is effected by the working of the regulating system and not — as in the case of the usual LC oscillators with a grid capacitor and grid leak — by the occurrence of grid current. The latter biases the control grid to such an extent that the valve is cut off during part of the cycle (class C operation) and the effective mutual conductance decreases. The anode current thus becomes pulse-shaped and consequently contains strong harmonics; these harmonics also occur in the circuit voltage, although to a lesser degree.

By using a regulating system, however, one can ensure that the valve remains operating in class A; thus operation is limited to a small, substantially straight part of the $i_a - v_g$ characteristic, resulting in much less distortion of the circuit voltage.

The last is not only important if a sinusoidal output voltage is desired, but also because distortion is one of the causes of frequency fluctuations. As Horton has shown²⁾, distortion results in a phase shift in the voltage which is fed back to the grid, so that the frequency f for which the oscillatory condition is complied with, is no longer the resonant frequency f_0 of the LC circuit. The difference between f and f_0 increases as the distortion becomes larger and will therefore not be constant if the supply voltage fluctuates, if the valve heats up, and if it ages in the long run.

Another cause of the frequency shift may lie in the well-known fact that the inter-electrode capacitances of a valve — especially the grid-to-cathode capacitance — depend on the space charge. In the circuit to be discussed the regulating system will keep the direct current which flows through the

¹⁾ See e.g. J. H. Roosdorp, On the regulation of industrial processes, Philips tech. Rev. Vol. 12, 221-227, 1951.

²⁾ J. W. Horton, Vacuum tube oscillators. A graphical method of analysis, Bell. Syst. tech. J. 3, 509-524, 1924.

valve approximately constant in case of mains voltage fluctuations and when the valve ages, so that the influence of this source of frequency fluctuations is reduced.

More detailed description of the regulating circuit

In various ways it is possible to effect the limitation of the amplitude, so that the oscillator remains operating in class A. In one well-known method a small incandescent lamp³⁾ (or another device having a resistance dependent on temperature) is used in one arm of a Wheatstone bridge; this bridge can consequently only be in equilibrium at one particular value of the input voltage.

Another method has been stated by Arguimbau⁴⁾, who uses a separate rectifier which converts an alternating voltage proportional to the oscillator voltage into a direct voltage, the latter serving as bias in the control grid circuit.

Taking the discussion of the regulating principle given above as a starting point, we can describe our method⁵⁾ as one in which two currents are compared instead of two voltages. Like Arguimbau we use an auxiliary rectifier; it consists of a diode D_1 and a blocking capacitor C_a (fig. 1a),

connected in series between the cathode and the anode of the oscillating pentode P . Consequently a pulsating voltage v_d is produced across the diode, with the anode negative with respect to the cathode (fig. 1b); the amplitude of v_d is $2V_0$ and its average value $E_0 = -V_0$. A capacitor C_g is connected between the points G and K (fig. 1a) of the grid circuit; G is connected to the anode of the diode via a resistor R_1 and to a point having a positive voltage; the reference voltage E_{ref} , via a resistor R_2 . Currents I_1 and I_2 flow through these resistors and with a negligible grid current the following formula applies:

$$I_1 + I_2 = 0. \dots \dots (1)$$

If E_{g1} is the direct voltage across the capacitor C_g , then $I_1 = (E_0 - E_{g1})/R_1 = -(V_0 + E_{g1})/R_1$ and $I_2 = (E_{ref} - E_{g1})/R_2$. Substitution of these values in (1) gives:

$$V_0 = \frac{R_1}{R_2} E_{ref} - \frac{R_1 + R_2}{R_2} E_{g1}. \dots (2)$$

E_{g1} — the control grid bias required to make the valve oscillate with the amplitude V_0 , at given supply voltages and circuit impedance — depends on the valve characteristics. It is, however, possible to make the first term in the right-hand member of (2), in which E_{g1} is absent, preponderate strongly with regard to the second term, for example by choosing E_{ref} high with respect to E_{g1} ; in first approximation we then find that

$$V_0 = \frac{R_1}{R_2} E_{ref},$$

viz. in first approximation the amplitude of the output voltage depends solely on the ratio of two fixed resistances and on one direct voltage, and consequently not on the valve characteristics and the circuit impedance. For a more accurate analysis, however, it is necessary to take into account also the influence of the last-mentioned quantities.

The stabilisation factor

The direct anode voltage, the screen-grid voltage and the heater voltage applied to the pentode, and also the reference voltage E_{ref} , are as a rule obtained from the A.C. mains and therefore show the same relative fluctuations as the mains voltage, unless special measures have been taken to prevent this.

It is obvious that, if V_0 is to be kept constant, E_{ref} should in any case be carefully stabilised, for example by means of one or more voltage reference tubes. Fig. 1a shows that the screen-grid voltage

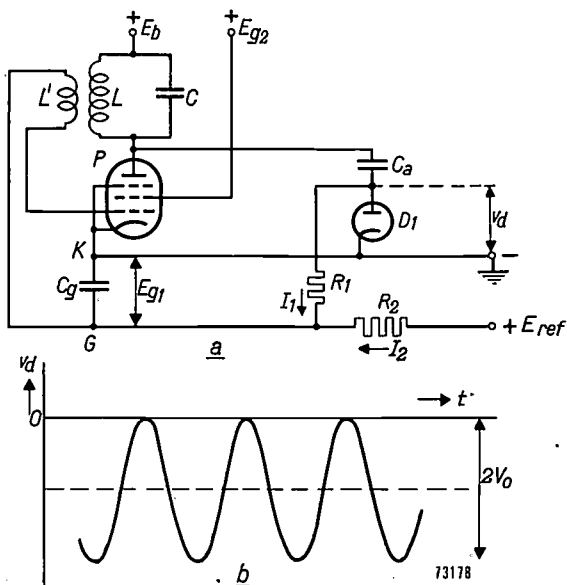


Fig. 1. a) Circuit for automatically stabilising the output voltage of an oscillator. P pentode. LC oscillatory circuit. L' feedback coil. C_a blocking capacitor and D_1 diode, rectifying the alternating anode voltage to a pulsating voltage v_d (represented in b as a function of the time t). E_{ref} reference voltage. C_g grid capacitor.

³⁾ L. A. Meacham, The bridge-stabilized oscillator, Proc. Inst. Rad. Engrs. 26, 1278-1294, 1938.
⁴⁾ L. B. Arguimbau, An oscillator having a linear operating characteristic, Proc. Inst. Rad. Engrs. 21, 14-28, 1933.
⁵⁾ This corresponds to a large extent with the triplex diode circuit used in some radio receivers for delayed automatic gain control; see e.g. J. Deketh, Fundamentals of radio-valve technique I, pp. 329-332 (Philips Technical Library, 1949).

can be obtained from the same voltage source without any difficulty; it is then no longer necessary to be concerned regarding the influence of fluctuations on the screen-grid voltage.

The exact value of the direct voltage in the anode circuit is not a factor of decisive influence; the properties of a pentode are sufficiently independent of this voltage. Heater voltage fluctuations, however, result in variations in the mutual conductance of the valve characteristic; they can be considered together with changes in mutual conductance due to other causes.

Finally, we have yet to consider the influence of changes in the impedance in the anode circuit.

If the mutual conductance S changes by an amount ΔS (due to a change in the heater current or by ageing of the valve), the original value is restored by changing E_{g1} by an amount

$$\Delta E_{g1} = -\frac{\partial E_{g1}}{\partial S} \Delta S,$$

in which $\partial E_{g1}/\partial S$ can be determined from the $S = f(E_{g1})$ characteristic. From equation (2) follows:

$$\Delta V_0 = -\frac{R_1 + R_2}{R_2} \Delta E_{g1} = \frac{R_1 + R_2}{R_2} \frac{\partial E_{g1}}{\partial S} \Delta S.$$

We now find for the ratio of the relative change of the mutual conductance to that of the amplitude, i.e. the stabilisation factor for mutual conductance changes:

$$a_s = \frac{\Delta S/S}{\Delta V_0/V_0} = \frac{V_0}{S} \frac{R_2}{R_1 + R_2} \frac{\partial S}{\partial E_{g1}} \quad (3)$$

The alternating voltage (amplitude V_0) across the impedance Z in the anode circuit is produced by the alternating anode current (amplitude $I_a = V_0/Z$), which in its turn is caused by an alternating voltage (amplitude V_{g1}) on the control grid: $I_a = S V_{g1}$. It will be assumed that this alternating grid voltage, which is due to the feedback, is a fraction t of the alternating anode voltage: $V_{g1} = t V_0$. To maintain the oscillatory state, the known oscillatory condition should be complied with:

$$SZt = -1, \quad (4)$$

which expresses that the total amplification for one round trip through the oscillator should be exactly equal to unity.

The impedance Z is variable in oscillators with adjustable frequency; moreover it varies with changes in the load on the oscillator and it can also vary due to incidental causes, for example by increase of the losses caused by dust and moisture. We see from (4) that changes in Z with constant t

are compensated by equal relative changes in S . Consequently the stabilisation factor for impedance variations,

$$a_z = \frac{\Delta Z/Z}{\Delta V_0/V_0},$$

is equal to the value of a_s found in (3).

For the sake of completeness we shall finally consider the case of a non-stabilised screen-grid voltage.

It follows from (2) that a change ΔE_{g1} of the control-grid bias results in a change of the amplitude V_0 by an amount ΔV_0 of the following magnitude:

$$\Delta V_0 = -\frac{R_1 + R_2}{R_2} \Delta E_{g1}.$$

For a pentode it holds that with constant anode current a change ΔE_{g2} in the screen-grid voltage is compensated by a change $\Delta E_{g1} = -\Delta E_{g2}/\mu_{g2g1}$ in the control-grid voltage; μ_{g2g1} is a factor larger than unity (e.g. 18 with a type EBF 80 pentode).

A change in the screen-grid voltage might be thought to influence the mutual conductance. However, this is barely so with a pentode; the characteristics of the anode current as a function of E_{g1} , with E_{g2} as parameter, run almost parallel and consequently the mutual conductance which is required for oscillating is reached at a certain value of the anode current which is substantially independent of E_{g2} .

Consequently $\Delta E_{g2}/\Delta V_0$ (the regulating factor) becomes

$$\frac{\Delta E_{g2}}{\Delta V_0} = \frac{R_2}{R_1 + R_2} \mu_{g2g1}.$$

For the ratio of the relative changes, the stabilisation factor for changes in screen-grid voltage, we find:

$$a_{E_{g2}} = \frac{\Delta E_{g2}/E_{g2}}{\Delta V_0/V_0} = \frac{V_0}{E_{g2}} \cdot \frac{R_2}{R_1 + R_2} \mu_{g1g2}.$$

Numerical example

In agreement with a circuit we will discuss presently, we have selected as a valve the EBF 80 pentode. Let V_0 be (say) 40 V and $Z = 30$ k Ω , then I_a is 1.33 mA. With $t = 1/20$, V_{g1} has to be 2 V and the required mutual conductance is then $S = 1.33/2$ mA/V = 0.67 mA/V.

Keeping the screen-grid voltage constant at a value of 85 V, the mutual conductance of 0.67 mA/V is obtained, according to the valve data, at a grid bias $E_{g1} = -4$ V, and amounts to $\partial S/\partial E_{g1} = 0.6$ mA/V² at this adjustment.

Choosing also 85 V for the reference voltage, and $R_1 = 1$ M Ω , then from (2) a value of 2.5 M Ω follows for R_2 .

According to (3) the stabilisation factor for mutual conductance and impedance changes now becomes:

$$|a_s| = |a_z| = \frac{40}{0.67 \times 10^{-3}} \cdot \frac{2.5}{1 + 2.5} \cdot 0.6 \times 10^{-3} = 26.$$

In this example the relative variations in the alternating voltage amplitude are consequently $1/26$ of the relative mutual conductance or impedance variations. This important improvement is obtained by means of only a few extra circuit elements. Considerably larger values of the stabilisation factor can be obtained by connecting an amplifier between the oscillatory circuit and the regulating device, as will be shown presently.

It should be observed that although stabilisation may be obtained to a certain degree with a conventional LC oscillator by stabilising the supply voltages (including the heater voltage), changes in mutual conductance by the ageing of the valve and impedance variations in the circuit may still occur. With the regulating circuit described above these influences are also kept in check.

Moreover, the regulating system keeps the valve constantly operating in class A. This not only offers the advantage that the distortion of the circuit voltage is very small — an advantage already mentioned above — but also that the interelectrode capacitances of the valve remain constant when the mains voltage fluctuates or the tube ages, since the anode current is kept constant in that case.

Adjustment and modulation of the alternating voltage amplitude

By making the reference voltage E_{ref} controllable, a simple means is obtained of varying V_0 or adjusting it to a given value. The graph showing V_0 as a function of E_{ref} in accordance with equation (2) is given in fig. 2.

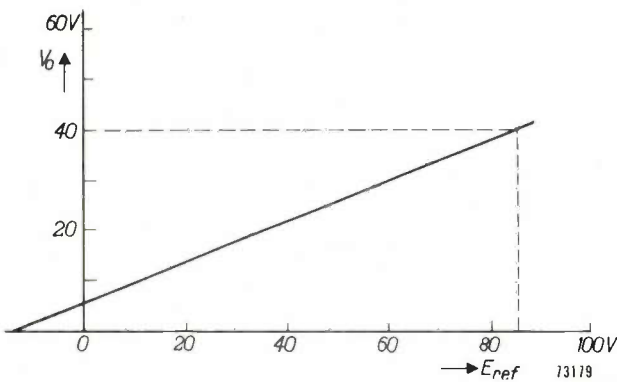


Fig. 2. Graphic representation of equation (2): amplitude V_0 of the alternating voltage versus reference voltage E_{ref} for $R_1 = 1 \text{ M}\Omega$, $R_2 = 2.5 \text{ M}\Omega$ and $E_{g1} = -4 \text{ V}$.

If E_{ref} is varied around a certain value, the amplitude of the output voltage is modulated. Use can be made of this if a modulated output voltage is required. Fig. 2 also represents the modulation characteristic, provided no grid current is flowing. If grid current does flow, equation (1) is no longer

applicable; the modulation characteristic then becomes less steep and consequently shows a sharp bend (fig. 3). Therefore the fact whether grid current flows or not, determines the greatest modulation depth at which no distortion occurs.

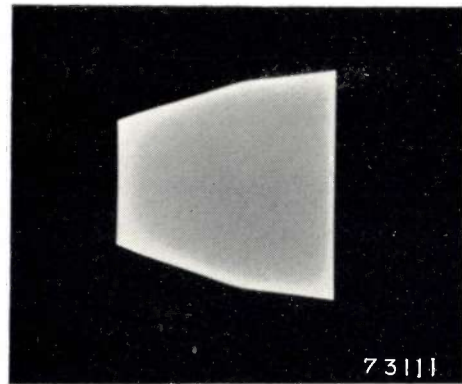


Fig. 3. Oscillogram of the alternating voltage of an oscillator in accordance with fig. 1a, which is being modulated with a sinusoidal voltage. The voltage for the horizontal deflection is likewise sinusoidal and in phase with the modulating voltage. For registering this oscillogram, the valve was so adjusted that grid current occurred, which caused a sharp bend in the envelope.

Plotting the modulation characteristic (or displaying the output voltage on an oscilloscope) is the simplest method of knowing whether grid current flows or not. The more usual method, i.e. to connect a voltmeter parallel to the grid resistor, would increase the stray capacitances to such an extent that the working point of the valve would be shifted considerably in order to remain in compliance with the oscillatory condition, so that the measurement would then be taken under very different conditions.

An upper limit is also set to the modulation frequency viz. by the tendency towards squegging.

Squegging and starting

The operation of the oscillator is determined mainly by the magnitudes of the capacitances C_g and C_a (fig. 1a). The capacitor C_g in the grid circuit, for example, together with the capacitance between control grid and earth, forms a voltage divider; if C_g is too small, a large proportion of the voltage induced in the feedback coil is lost in C_g . On the other hand, C_g should not be too large either — and this also applies to C_a — since the control voltage is then built up so slowly that there is a risk of squegging; the normal oscillating process is then periodically interrupted at a frequency f_{rel} , depending on a relaxation time.

It is possible to investigate whether there is a risk of squegging by modulating the oscillator in the above-mentioned way with an alternating voltage of constant amplitude and variable fre-

quency and by displaying the modulated output voltage on an oscilloscope. If the modulation frequency approaches f_{rel} , a deepening of the modulation is observed. This puts an upper limit to the modulation frequency. A square-wave modulating voltage can also be used; the tendency towards squegging can in that case be judged from the damping of the oscillations present in the envelope of the modulated output voltage wave (fig. 4).

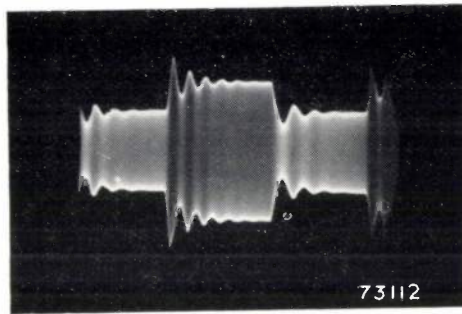


Fig. 4. Oscillogram of the voltage across the LC circuit of an oscillator in accordance with fig. 1a, modulated with a square-wave voltage. The tendency towards squegging can be judged from the damping of the oscillations present in the envelope.

If C_g and C_a have been so chosen that the oscillator (connected in accordance with fig. 1a) can

the resistor R_1 and the diode D_1 , while the greater part flows through the feedback coil and further as grid current through the pentode. Due to the last fact the path between the control grid and cathode has a differential resistance which is much smaller than the impedance of C_g ; consequently only a small part of the growing alternating voltage which is induced in the feedback coil reaches the control grid; moreover it does not have the correct phase. For this reason the oscillating process cannot start.

To remove this obstacle, provision should be made that I_2 cannot flow as grid current. To attain this, a diode D_2 can be incorporated (fig. 5a), offering an easy path to I_2 , and a resistor R_3 , increasing the resistance in the path via the control grid.

Fig. 5b shows that a single valve, containing a pentode system and two diode systems, such as the above-mentioned EBF 80, can be used to advantage. The reference voltage and the screen-grid voltage are stabilised at 85 V by one 85 A1 voltage reference tube ⁶⁾. Consideration is being given to the application of a system similar to that just mentioned in various measuring apparatus brought on the market by Philips.

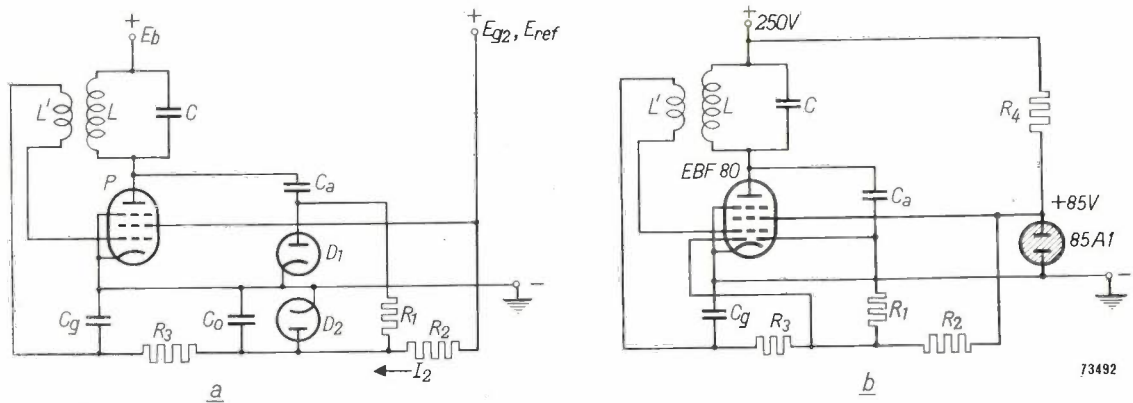


Fig. 5. a) As fig. 1a, but with the addition of a diode D_2 and a resistor R_3 , which prevent the current I_2 from flowing across the control grid immediately after switching on; this would hamper the oscillator from starting properly. C_0 smoothing capacitor. Meaning of the other letter references as in fig. 1a. b) Practical example of the principle with an EBF 80 valve, in which the pentode P and the diodes D_1 and D_2 are combined. The reference voltage, also screen-grid voltage, is kept constant by a 85A1 voltage reference tube, with the series resistor R_4 .

oscillate without tendency towards squegging, it apparently has some difficulty in starting.

The cause of this is the following: Immediately after switching on, the valve by no means oscillates at full strength and the alternating anode current is still negligible; the current I_2 , due to the reference voltage source, now flows for a small part through

Installation for the calibration of valve voltmeters

The stabilisation of an oscillator voltage has already found practical application in an installation for the calibration of valve voltmeters.

Fig. 6 shows this installation for the calibration

⁶⁾ T. Jurriaanse, A voltage stabilizing tube for very constant voltage, Philips tech. Rev. 8, 272-277, 1946.

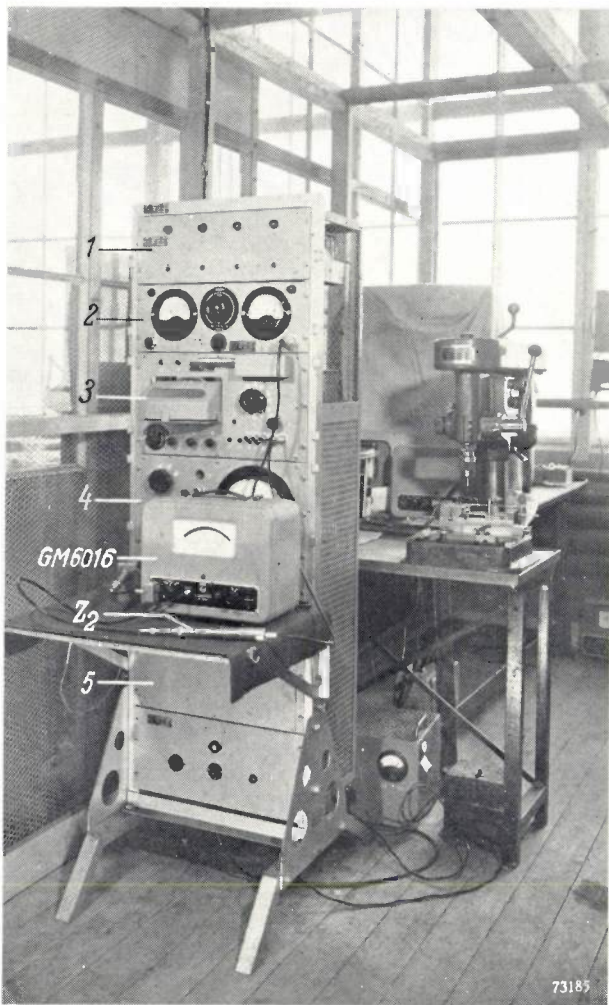


Fig. 6. Calibrating rack for the electronic millivoltmeter type GM 6016. 1 panel with four stabilised supply rectifiers. 2 panel for the adjustment of the supply voltage (50 c/s) for the meter to be calibrated. 3 panel with calibration voltages from 3 mV to 300 V, 100 kc/s, for the calibration of meter GM 6016 and the accessory attenuator (Z_2). 4 panel for checking the frequency characteristic. 5 panel with twelve oscillators (frequencies 1 kc/s to 30 Mc/s), with automatically stabilised output voltage. To the right, the drilling machine for drilling the catch holes in the calibrated attenuator.

of millivoltmeters type GM 6016, an instrument suitable for frequencies ranging from 1000 c/s to 30 Mc/s. The input voltage required to fully deflect the meter amounts to 3 mV at maximum sensitivity⁷); this voltage can be increased in steps up to 1000 V by means of a capacitive attenuator.

A similar installation is in use for the calibration of millivoltmeter type GM 6005; this is a meter suitable for frequencies from 20 c/s to 1 Mc/s which gives full deflection at voltages from 10 mV to 300 V. In the following we shall only discuss the calibration installation for type GM 6016.

⁷) It is only with regard to sensitivity that type GM 6016 differs from type GM 6006 described before, which already gives full deflection at 1 mV (H. J. Lindenhovius, G. Arbelet and J. C. van der Breggen, Philips tech. Rev. 11, 206-214 1950.

Two checks can be carried out with the installation under discussion :

- 1) absolute calibration of the meter including the attenuator, and
- 2) checking of the frequency characteristic.

An oscillator O_0 for 100 kc/s (fig. 7) is mounted in panel 3 (fig. 6) for the first-mentioned calibration; it is connected to an amplifier A_0 , supplying a voltage with an r.m.s. value of 300 V. This voltage is stabilised by a regulating device R as discussed above; the components C_a , D_1 , C_g , R_1 and R_2 correspond with those correspondingly marked in fig. 1a. The only difference between the two circuits is the presence of an amplifier in the one under discussion, which makes no difference to the principle, but as observed before, improves the stabilisation factor. The reference voltage E_{ref} is obtained from a stabilised direct voltage source of 310 V via a potentiometer.

Panel 5 (fig. 6) contains twelve oscillators, for the frequencies 1, 5, 100 and 1000 kc/s and 1, 3, 10, 15, 18, 20, 25 and 30 Mc/s, for checking the frequency characteristic⁸). The circuit voltage of these oscillators is stabilised at 30 V (r.m.s. value), by application of the regulating principle discussed above.

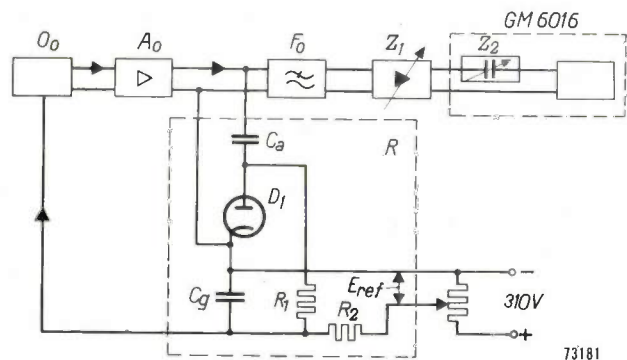


Fig. 7. Circuit diagram for the calibration of the meter GM 6016 and its attenuator Z_2 . O_0 oscillator with frequency 100 kc/s. A_0 amplifier. R , with the component parts C_a , D_1 , C_g , R_1 and R_2 ; unit for the producing of a control voltage (see fig. 1a) which stabilizes the output voltage of A_0 . The reference voltage E_{ref} is adjustable. F_0 low-pass filter. Z_1 attenuator adjustable in calibrated steps.

Absolute calibration and checking of the attenuator at 100 kc/s

Although it actually falls outside the scope of this article, a brief outline of the calibration procedure is given below.

An attenuator Z_1 (fig. 7) belonging to the calibrating installation and to which 300 V at 100 kc/s is applied, delivers calibrating voltages of 300, 100, 30 and 10 V to a series of sockets, and in addition to this, by means of a switch, voltages of 3 and 1 V and 300, 100, 30, 10 and 3 mV can be obtained.

⁸) The presence of two oscillators for 1000 kc/s (= 1Mc/s) will be explained later.

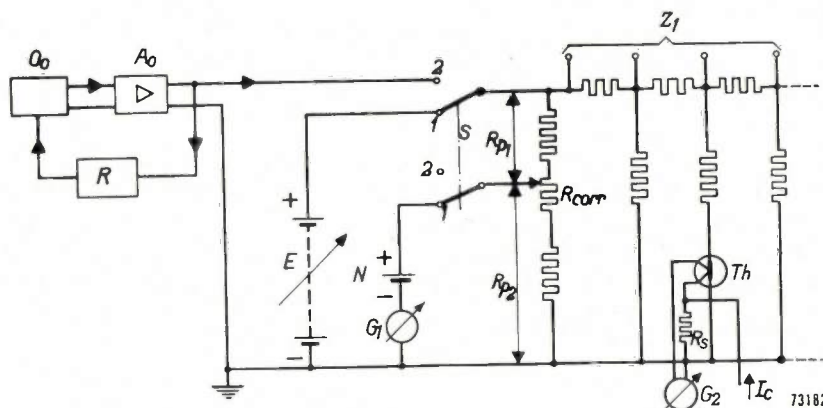


Fig. 8. Device for adjusting the 100 kc/s input voltage of attenuator Z_1 exactly to 300 V. Switch S in position 1: After the correction potentiometer R_{corr} has been set in accordance with the temperature of the standard cell N_1 , the direct voltage E is so adjusted that the deflection of the galvanometer G_1 becomes zero, E is then exactly 300 V. The deflection of the galvanometer G_2 , which is connected to four thermocouples Th connected in series (of which only one is shown in the diagram), is reduced to zero by means of the current I_c , flowing through the resistor R_s . Switch S in position 2: The reference voltage E_{ref} of the oscillator-amplifier O_0-A_0-R (see fig. 7) is so adjusted that G_2 again returns to zero at exactly the same value of I_c as before. The r.m.s. value of the voltage at the input of Z_1 is then exactly 300 V.

First the calibrating voltage of 3 mV is applied direct to the terminals of the meter GM 6016 which has to be checked; the meter is so corrected that it gives the proper deflection. Next one changes over to the higher calibrating voltages, the capacitive attenuator belonging to the meter (Z_2 , fig. 7) being interconnected. This attenuator, which is continuously variable, is each time adjusted in such a way that the meter gives the correct indication; at these positions holes are drilled, so that a catching device, consisting of a steel ball, which is pressed into the hole by a spring, stops the attenuator.

The question, however, arises how it can be ascertained that the high-frequency alternating voltage at the input of Z_1 is exactly 300 V. This is checked by means of a direct voltage, by comparison with the voltage of a standard cell. The procedure is as follows.

The input side of Z_1 is disconnected from the amplifier A_0 and connected to an adjustable direct voltage source $E \approx 300$ V (fig. 8). To make E exactly 300 V it would of course be possible to use a moving-coil voltmeter, but greater accuracy can be obtained by comparing a given fraction of E with the voltage of a standard cell in a compensation circuit. Therefore, the direct voltage source is shunted by a potentiometer, consisting of the (substantially) fixed resistors R_{p1} and R_{p2} (fig. 8). The ratio of these resistances has been so selected that the voltage across R_{p2} is equal to that of the standard cell N_1 if E is exactly 300 V. To make $E = 300$ V, E is varied until the deflection of the galvanometer G_1 has become zero.

Since the voltage of a standard cell depends to a certain degree on the temperature, the ratio $R_{p1} : R_{p2}$ has to be variable to a slight extent; the small correction potentiometer R_{corr} serves this purpose. This potentiometer is provided with a temperature scale; it is set to the temperature read on a thermometer inserted in the box of the standard cell (Philips type GM 4569, fig. 9).

If a direct voltage of exactly 300 V at the input of the attenuator Z_1 has thus been obtained, a certain current will flow through each of the resistors of Z_1 . One of these currents is measured in relative value by means of four thermocouples Th connected in series (four to increase the sensitivity)

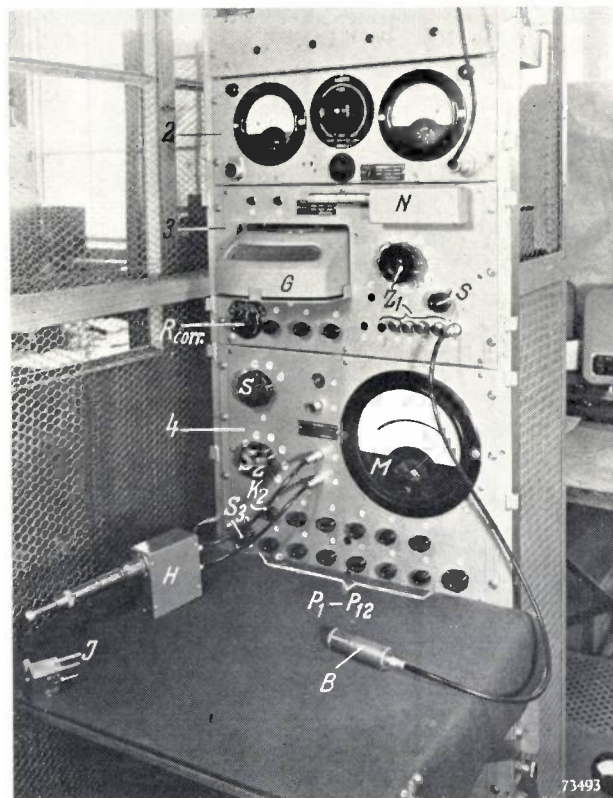


Fig. 9. Close-up view of the panels 2, 3 and 4. Cf. fig. 8.: N standard cell (type GM 4569) with thermometer. G galvanometer serving either as G_1 or as G_2 . S switch changing over from direct voltage to alternating voltage. Z_1 attenuator adjustable in steps. B socket to which the attenuator of the meter GM 6016 is plugged. Cf. fig. 10: S_1, S_2, S_3 frequency selector switches. P_1-P_{12} potentiometers for adjusting the oscillator voltages to the correct value. K_2 coaxial cable. H probe with germanium diode and capacitive voltage divider (G_2 and C_1-C_2 in fig. 10). The meter GM 6016 to be checked is connected to the plug J . M meter for the comparing the oscillator voltages.

and a galvanometer G_2 (fig. 8). This too, is effected by a zero method, for the galvanometer circuit includes a resistor R_s which is traversed by an adjustable auxiliary direct current I_c . This current is adjusted in such a way that the deflection of the galvanometer G_2 becomes zero, i.e. that the voltage across R_s exactly compensates the output voltage of the thermocouples; I_c then is a measure of the current flowing through the heaters of the thermocouples. Now Z_1 is switched over from the direct voltage source to the amplifier A_0 , which supplies an alternating voltage of 100 kc/s, and the reference voltage E_{ref} is adjusted until the deflection of the galvanometer G_2 again becomes zero. In that case the alternating voltage at the input of Z_1 has an r.m.s. value of exactly 300 V (0.1% difference already makes G_2 deflect 15 gradations), and at the output terminals of Z_1 the calibrating voltages 300 V, 100 V, 30V, ...3 mV are available. The attenuator Z_1 has been corrected for stray capacitances, so that the attenuation is the same with an alternating voltage of 100 kc/s as with a direct voltage.

The deflection of the meter GM 6016 depends on the form factor of the input voltage; the meter, however, has been calibrated in r.m.s. values. The calibration is correct only for sinusoidal voltages. Although the output voltage of the amplifier A_0 shows only little distortion, a simple low-pass filter F_0 has been placed behind this amplifier; in this way the distortion of the voltage across the meter GM 6016 is kept lower than 0.1%.

Checking the frequency characteristic

Via a coupling coil a voltage of 300 mV is obtained from each of the twelve oscillators of panel 5. These coupling coils are connected to two selector switches (S_1 and S_2 , fig. 10) on panel 4 by coaxial cables; with these switches a measuring

provided with a sliding core of Ferroxcube; see the article referred to in note ⁷), page 209).

Each of the twelve oscillators is provided with a potentiometer ($P_1 \dots P_{12}$, fig. 9) with which the reference voltage and consequently the output voltages can be varied.

The way in which the equality of the measuring voltages is checked, requires further comment. A direct voltage is derived from the alternating voltage by means of a germanium diode Ge_1 . A vibrating contact (S_4 , fig. 10) converts this into a low-frequency pulsating voltage which is amplified by the amplifier A_1 and measured in a relative measure by the meter M . At the lower frequencies (1 to 1000 kc/s), from which a choice can be made by means of the selector switch S_1 , a cathode follower is incorporated behind S_1 . The germanium diode is connected direct to this cathode follower, and the meter GM 6016, which has to be checked, via a coaxial cable. However, at frequencies exceeding 1 Mc/s the difference between the voltages at the beginning and the end of the cable would become excessive (see the article mentioned in note ⁷), page 214); therefore, at these higher frequencies, use is made of a germanium diode Ge_2 mounted in a probe at the same end of the cable where the meter is connected.

The frequency 1000 kc/s = 1 Mc/s occurs both in the lower and in the higher frequency range, so as to ensure a good agreement between the measurements in the two ranges.

By means of an attenuator a voltage of 3 mV is derived from the voltage of 300 mV which is induced in each of the coupling coils, for checking the meter without using the attenuator Z_2 belonging to it. For the frequencies from 1 to 1000 kc/s the first-mentioned attenuator consists of resistors (r_1, r_2 , fig. 10), which are contained in the measuring rack; for the frequencies 1 to 30 Mc/s it consists of capacitors (C_1, C_2), which are mounted in the probe at the end of cable K_2 .

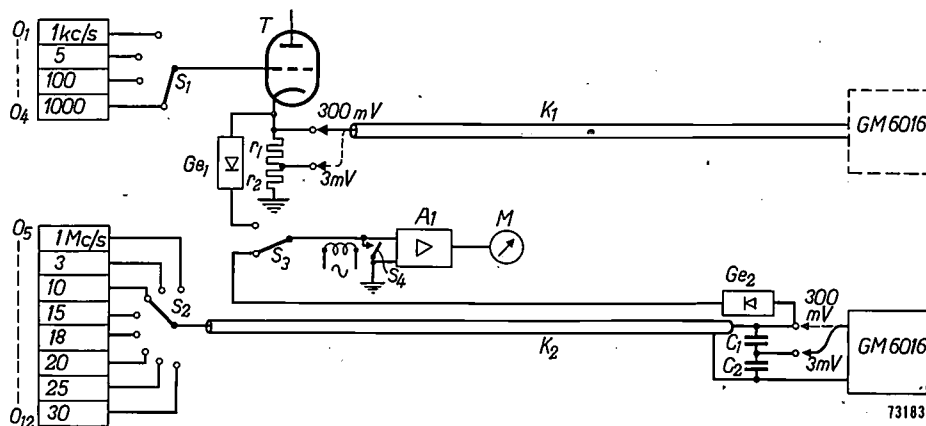


Fig. 10. Set-up for checking the frequency characteristic of the millivoltmeter GM 6016. O_1 - O_{12} oscillators with automatically stabilised voltages, with frequencies 1 kc/s to 30 Mc/s. S_1 selector switch 1-1000 kc/s. S_2 selector switch 1-30 Mc/s. S_3 selector switch (1-1000 kc/s)/(1-30 Mc/s). T cathode follower. r_1, r_2 and C_1, C_2 voltage dividers, by which not only 300 mV but also 3 mV are available. K_1, K_2 , coaxial cables. Ge_1, Ge_2 rectifying circuits with germanium diode (in the form of a probe). S_4 vibrating contact. A_1 amplifier. M meter.

voltage in either frequency range 1-1000 kc/s or 1-30 Mc/s is selected. With equal measuring voltages, the meter GM 6016 should give the same deflection within certain tolerances; should it fail to do so, then certain components of the meter should be readjusted until the frequency characteristic has become sufficiently flat (to facilitate this, certain coils are

Summary: In accordance with a well-known principle of regulating technique, the output voltage of a valve oscillator may be stabilised by the mutual comparison of two currents, one of which is obtained by rectifying the output voltage, whereas the other originates from a reference voltage source; from the difference between these currents a regulating voltage is derived, which serves for biasing the oscillator valve.

In this way, the magnitude of the output voltage can, with the aid of only a few extra circuit elements, be made highly insensitive to mains voltage fluctuations and to changes in the mutual conductance of the valve and of the impedance of the oscillatory circuit. The output voltage is less distorted than with an ordinary LC oscillator.

By making the reference voltage controllable, it becomes possible to adjust the output voltage to the value required, and by superimposing an alternating voltage on the reference voltage, the amplitude of the output voltage can be modulated. The upper limit of the modulation frequency and that of the modulation depth are discussed.

The tendency towards squegging can be counteracted by

giving the correct value to certain capacitances. For the oscillator to start smoothly after switching on, care should be taken to prevent the current supplied by the reference voltage source from flowing as control grid current through the valve; to this end an auxiliary diode can be inserted in the circuit.

The valve EBF 80 contains a pentode system, which may serve as an oscillator valve, and two diode systems, one of which can rectify the alternating voltage to be stabilized, while the other may function as the auxiliary diode just mentioned.

A calibrating system for the type GM 6016 electronic millivoltmeter (1000 c/s to 30 Mc/s) is discussed as a practical application.

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 ~~XXXXXXXXXXXXXXXXXXXX~~ Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

2012: R. van der Veen: Fluorescence and induction phenomena in photosynthesis (Physiologia plantarum 4, 486-494, 1951).

Fluorescence of chlorophyll in leaves shows induction phenomena during the first few minutes of illumination, which are very much like those of photosynthesis. The influence of preceding dark period, of temperature, of light intensity and of CO₂ and O₂ pressure on the adaptation curve was studied. An increase of photosynthesis during adaptation is nearly always correlated with a decrease of fluorescence.

The only exceptions are: (a) in the temperature limited range of photosynthesis, a lower temperature causes less photosynthesis, but does not result in any marked change in fluorescence and b) a high O₂ pressure causes both a lower rate of photosynthesis and lower fluorescence.

The results of these experiments are therefore mostly in agreement with the view of Katz; the fluorescence of chlorophyll can often be used as a flow meter for the energy of photosynthesis. Sometimes, however, photosynthesis and fluorescence do not show any correlation. In these cases it is probable that the light energy is used for the splitting of H₂O but that afterwards back reactions cause a reduction of the photosynthesis.

2013: W. de Groot: On the definition of standard illuminant A (Physica 17, 920-922, 1951, No. 10).

In the Stockholm meeting 1951, the C. I. E. provisionally defined standard illuminant A by the

value $\beta = c_2/T = 5.0386 (= 14350/2848 \approx 14380/2854)$ to be used in Planck's (or Wien's) radiation formula. It is pointed out that if one wishes to realise the standard (in Wien's approximation) by constructing a black body, the monochromatic power ratio of which (with respect to a black body at the melting point of gold) is $\exp(a/\lambda)$, where $a = c_2/T_{Au} - \beta$, a change in c_2 and T_{Au} in general influences a , if β is kept constant. The spectral distribution of this source will in general deviate from the calculated value unless the "true" values of c_2 and T_{Au} are used in calculating a .

2014: A. Bierman: A systematic method with new standards to determine the correct magnitude of radiographical exposures (Acta radiologica 36, 311-323, Oct. 1951).

The value of a radiographical exposure E is determined by $E = (kV/100)^5 \text{ mA} \cdot \text{sec}/\text{m}^2$ (in what the writer calls "radiographic exposure units" or "re u"). A standard object (a water phantom with definite dimensions) is used to characterise the radiographical outfit with the "specific exposure" E_s . This is the value of E required to obtain a film density of unity when radiographing the standard object. The exposure required for a certain object can be computed with the aid of E_s and "relative exposure" E/E_s . The values of E/E_s appropriate to the various objects constitute a "basic exposure table", from which the correct exposure is derived by means of the formula $E = (E/E_s) \times E_s$. All computations can be avoided by the use of two nomograms.

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 DECADE COUNTER TUBE FOR HIGH COUNTING RATES

by A. J. W. M. VAN OVERBEEK, J. L. H. JONKER and K. RODENHUIS.

621.385.832: 621.318.57

Mechanical, electromagnetically operated counting devices which count the applied number of electrical pulses, have been known for a long time. In recent years there has been an ever increasing demand for counting devices which operate at a much higher speed. This demand exists both in telecommunication and for modern computers, for measuring frequencies and for carrying out nuclear measurements. It is only by electronic means that a solution is offered in these cases.

A new electronic tube, the EIT, which has been specially designed for counting purposes, can count pulses at a very high rate. The number of counts can be read directly on the tubes themselves.

There are various types of mechanical counting devices which are operated by an electromagnet under the influence of electrical pulses. With some of these the number of pulses can be read in figures, such as with counters for telephone calls; with others a switch lever is moved a number of steps equal to the number of pulses received, so that a given connection is established, as in the case with selectors in a telephone exchange. Due to the delay in build-up and die-away of a magnetic field and the inertia of the moving parts, the counting rate of such electromechanical counters is not very high. For computers which must make thousands of additions per second, for the pulse code systems used in telecommunication, etc., this rate is quite insufficient.

Electronic tubes have, of course, been widely used for counting purposes. With valves of the conventional type, however, fairly large numbers of valves are required. A further problem is to render the counted number visible.

Reports on the development work carried out by Philips in the field of electronic valves designed for switching have been published both in this review ¹⁾ and elsewhere ²⁾. One of the tubes which has resulted from this research work is the decade counter tube. This tube has already been mentioned in the publications referred to in footnote ²⁾;

a description of this tube is now given and a new circuit is discussed.

Description of the decade counter tube

The counter tube is provided with an electron gun which emits an electron beam towards the envelope of the tube. A deflecting device with a feedback system, which is to be discussed later, ensures that the beam can occupy ten discrete positions. The envelope is lined with fluorescent material which fluoresces at a different place for each of the positions, so that the position of the beam can be read from the figures 0, 1, 2,.....9 placed opposite the fluorescing marks (*fig. 1*). Before counting starts, the beam occupies position 0, after the first pulse it is advanced to position 1, after the second pulse to position 2, etc., until position 9 is reached. At the tenth pulse the beam is reset to position 0, and at the same time a pulse is fed to a second counter tube, as a result of which its beam is moved from position 0 to position 1.

- ¹⁾ J. L. H. Jonker and Z. van Gelder, New electronic tubes employed as switches in communication engineering, Philips tech. Rev. 13, 49-54 and 82-89, 1951/52 (Nos 3 and 4).
- ²⁾ 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. J. L. H. Jonker, A. J. W. M. van Overbeek and P. H. de Beurs, A decade counter valve for high counting rates, Philips Res. Rep. 7, 81-111, 1952 (No. 2).

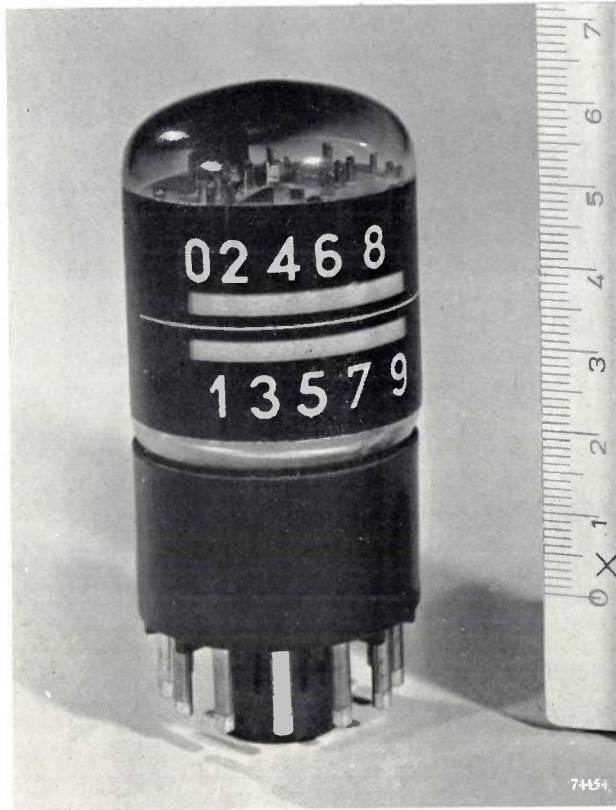


Fig. 1. The EIT decade counter tube. A blue fluorescent mark indicates the figure which corresponds to the position of the electron beam in the counter tube, i.e. to the number of pulses counted.

This second tube therefore counts the decades. A third tube can be added which counts the hundreds, a fourth which counts the thousands, etc. (fig. 2).

With n tubes in cascade it is thus possible to count up to $10^n - 1$ ($\approx 10^n$) pulses.

Fig. 3a shows a cross-section of the decade counter tube, whilst fig. 3b shows the diagrammatic representation which will be used in circuit diagrams. The cathode is of the conventional type as used in receiving valves, apart from the fact that it emits electrons from one side only of its rectangular cross-section. A control grid (g_1), four rod-shaped focusing electrodes (p_1, p_2) and an accelerating electrode (g_2), together with the cathode, form the electron gun. This has been so designed that the cross-section of the electron beam thus obtained is not circular but ribbon-shaped. As shown elsewhere^{1,2}, a ribbon-shaped electron beam has the following advantages:

- 1) It is easy to obtain a fairly strong beam current at a relatively low voltage, for example 1 mA at 300 V.
- 2) The dimensions of the tube can be kept small, roughly as those of a normal receiving valve. The EIT counter tube (fig. 1) has a maximum diameter of 37 mm and a maximum height of 83 mm.
- 3) It is necessary to align the electrodes, which are fixed between two mica discs as in the technique used for manufacturing receiving valves, in one dimension only, namely in the direction normal to the plane of the beam; with a beam having a circular cross-section it would on the other hand be necessary to align the electrodes in two directions, which would render manufacture much more difficult.

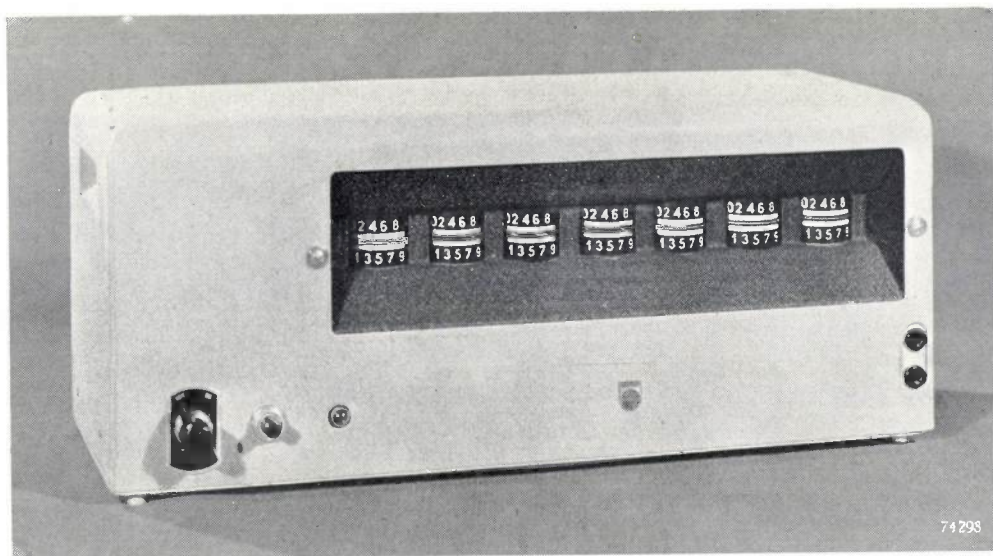


Fig. 2. Decade counter with seven decade counter tubes EIT connected in cascade, suitable for counting up to $(10^7 - 1)$ at a speed of 30,000 pulses per second. From left to right the millions, the hundred thousands, the ten thousands, etc. can be read on the tubes themselves.

Bottom left: mains switch, push-button switch with which all counter tubes can be reset to zero before counting starts, and pilot lamp.

Bottom right: input terminals.

The importance of the advantages mentioned under (1) and (2) can be stated as follows. The advantage of a strong current in the case of a counter tube is related to the counting rate. Generally speaking, the time required for an "electronic event" is proportional to VC/I , where V is the potential difference traversed by the electrons, I the current intensity and C the inter-electrode

capacitance. In order to keep this time short, it is thus favourable that V is kept small, I being given a large value. The ribbon shape of the beam is advantageous in both respects. Since C is proportional to the linear dimensions of the electrode system, attempts must moreover be made to keep this system small, i.e. to apply the technique used for receiving valves. It will be clear that the resulting small external dimensions of the tube are welcome also with a view to saving space.

The ribbon-shaped beam thus obtained proceeds between two deflection electrodes (D, D' , fig. 3). These have been so positioned that the beam almost touches one of these electrodes in each of its extreme positions, deflection sensitivity thus being at a maximum.

At given values of the deflection, the beam passes through one of the ten vertical slots in the slotted electrode (g_4), which is at a positive voltage. It will be shown later how a special circuit ensures that the beam can occupy only one of these ten positions. Some of the electrons passing through a slot impinge on the anode (a_2) placed behind the slotted electrode; the remainder pass through an aperture in this anode and impinge on the envelope. The part of the envelope situated behind the anode is lined with fluorescent material, so that a fluorescing mark is produced behind the opening in the anode through which electrons pass. The number (fig. 1) opposite this mark thus indicates the position of the beam.

When the beam passes through slot 9, it has not yet reached its extreme position, but can be deflected slightly further under the influence of a following pulse. When this occurs, it impinges on the so-called reset anode (a_1). The impinging of the beam on the reset anode initiates the resetting to position 0. Before discussing this, it will be shown how the beam is fixed at well-defined positions.

Step-wise deflection of the beam

Fig. 4a gives a schematic representation of an imaginary cathode-ray tube, containing an electron gun, a set of deflection electrodes and an anode. When the potential v_D of the one deflection electrode is kept constant and the potential $v_{D'}$ of the other is varied, the electron beam will move along the anode, but the anode current i_a remains constant (line I in fig. 4b). When the deflection electrode D' is connected to the anode and these two electrodes are fed via a common resistor R_a from a battery with voltage V_B (fig. 4c), the following equation, however, applies:

$$v_{a2} = v_{D'} = V_B - i_a R_a \dots \dots (1)$$

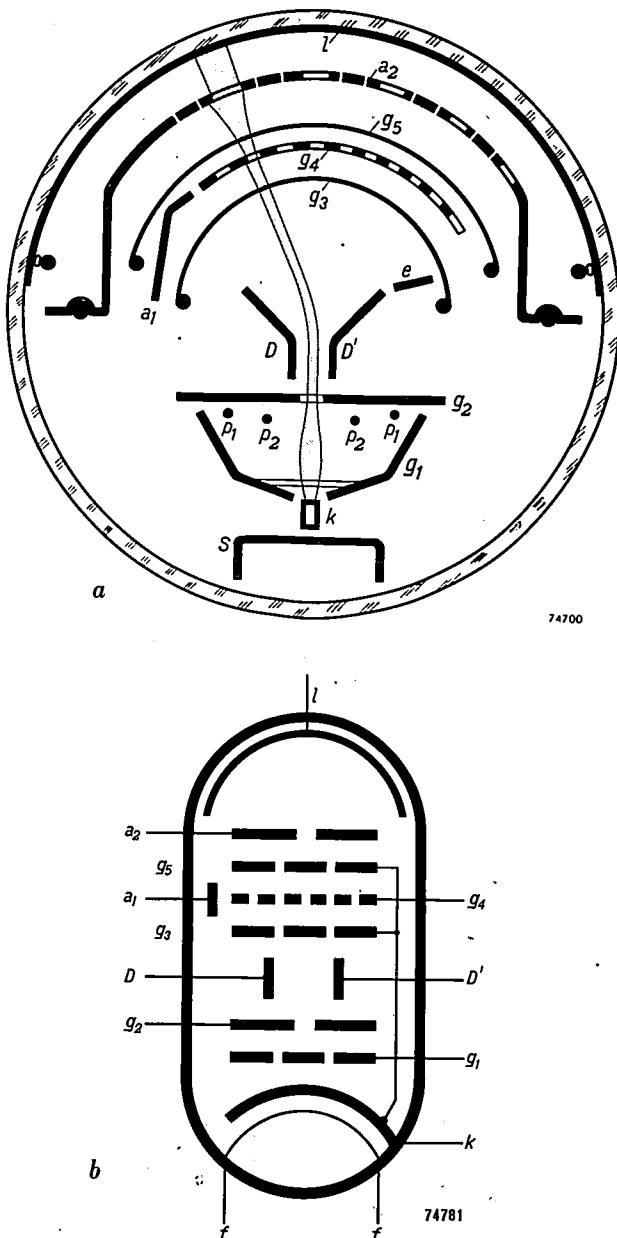


Fig. 3. (a) Cross section and (b) diagrammatic representation of the decade counter tube. The cathode k (with heater f), the control grid g_1 , the four internally connected focusing electrodes p_1 and p_2 and the accelerating electrode g_2 form the electron gun, which produces a ribbon-shaped electron beam (the width of the ribbon is normal to the plane of the drawing). D, D' deflection electrodes. g_3, g_5 suppressor grids. a_1 reset anode. g_4 electrode with ten slots. a_2 anode. l fluorescent layer applied to a conductive layer. Screen s (internally connected to k) prevents primary electrons from impinging on the envelope. Auxiliary anode e (internally connected to g_2) captures secondary electrons.

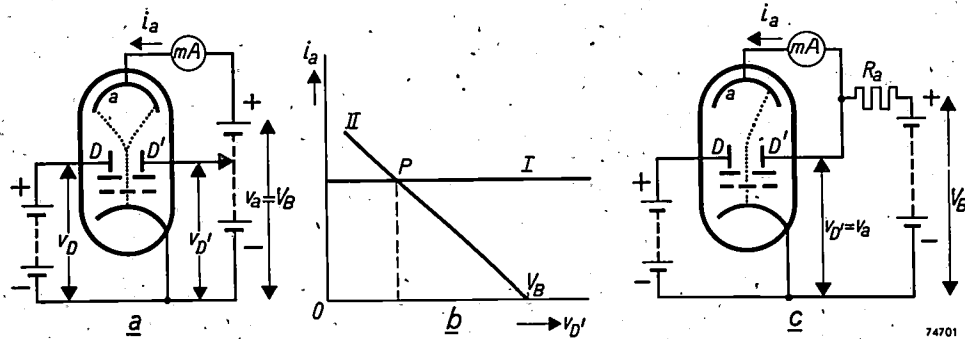


Fig. 4. (a) Imaginary cathode-ray tube with an electron gun, deflection electrodes D, D' and an anode a . At v_a and $v_D = \text{constant}$, the electron beam can be given any position between the extreme positions (dotted lines) by varying $v_{D'}$. In doing so the anode current i_a remains constant, see I in (b).
 (c) When D' and a are connected to the positive terminal of a battery with voltage V_B via a common resistor R_a , eq. (1) is applicable (represented by the line II in (b)). The beam can now occupy one position only, determined by the abscissa of the point of intersection P .

This relation is represented by the line II in fig. 4b. The point of intersection P of the lines I and II gives the state of equilibrium; the electrode D' then assumes a potential which corresponds to the abscissa of P , and according to this given potential the beam occupies a well-defined position.

Advantage is taken of this principle for fixing the positions of the beam in the counter tube. In this case too, the anode a_2 and the deflection electrode D' (i.e. the one nearest to the figure 0) are connected to a direct voltage source V_B via a common resistor R_{a_2} (fig. 5); the potential of D and a_2 is denoted by v_{D',a_2} . The straight line I of fig. 4b is now, however, replaced by the undulated curve I of fig. 6; the way in which this is obtained will be shown presently. This curve I is intersected 19 times by the resistance line II . Each of these points of intersection corresponds to a condition of equilibrium — although only the ten points of intersection numbered from 0 to 9 represent stable

positions. From this it may be seen that the position of the beam indicated diagrammatically in fig. 3a is one of its stable positions, i.e., passing partly through a slot and impinging partly on the

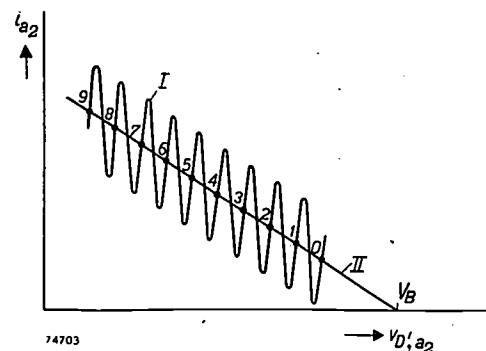


Fig. 6. The $i_{a_2} = f(v_{D',a_2})$ characteristic of the decade counter tube has the form of the undulated curve I . The straight line II represents the resistance line according to eq. (1). Only the points of intersection of I and II numbered from 0 to 9 represent stable positions of the beam; the other points of intersection correspond to unstable positions. V_B is only slightly larger than the abscissa of the point of intersection 0.

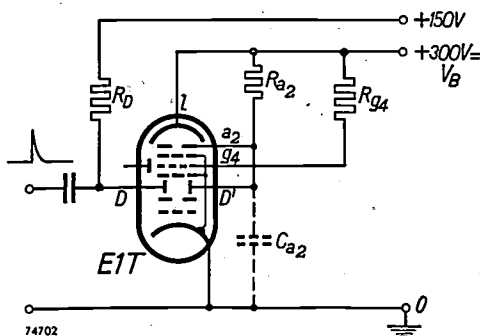


Fig. 5. EIT counter tube of which the anode a_2 and the deflection electrode D' are fed via a common resistor R_{a_2} , similar to the tube shown in fig. 4c. The beam has now the same number of stable positions as the number of slots in the electrode g_4 (see also fig. 6). By applying positive pulses to D , the beam can be advanced in steps (see later). C_{a_2} represents the stray capacitance.

metal at the left of this slot: in fact, when the beam is deflected further to the left the current i_a initially decreases, in accordance with the characteristic I near the points of intersection 0...9 (fig. 6).

How the characteristic $i_{a_2} = f(v_{D',a_2})$ is obtained

In fig. 7a a slotted electrode and an anode have been drawn as flat planes. It is assumed that the ten slots have the same dimensions and are equidistant and that the thickness of the ribbon-shaped beam exceeds the width of the slots and is less than the width of the spaces between the slots.

When the beam is now made to move from slot 0 to slot 9 by (continuously) lowering the voltage v_{D',a_2} , it might be expected at first sight that a characteristic $i_a = f(v_{D',a_2})$ as shown by curve I

of fig. 7b is obtained: each time the beam is directed on a slot, i_{a2} reaches a maximum (which has the same value for all slots), and each time the beam points to the centre between the slots, i_{a2} drops back to zero.

Assuming for the time being that the characteristic had this form, it would be possible to proceed according to fig. 5, and to give V_B and R_{a2} such values that the line II, which represents eq.(1) graphically, intersects all waves of I (fig. 7b). In practice, however, with a slotted electrode according to fig. 7a, the characteristic $i_a = f(v_{D',a_2})$ will not assume the form I of fig. 7b, but that of fig. 7c. The disadvantages of the latter form are that the magnitude of R_{a2} (the slope of line II) must remain within very narrow limits and that V_B must be given a much higher value to prevent one or several points of intersection from being obliterated at the left-hand side. The form of I as shown in fig. 7c is a result of the asymmetry in the deflection (v_D constant, $v_{D'}$ variable).

It is in fact due to this asymmetry that the focusing of the beam on the slotted electrode deteriorates and that the beam thus becomes wider as it is displaced further to the left. Consequently, the number of electrons per second passing through the slots having high numbers is smaller than that

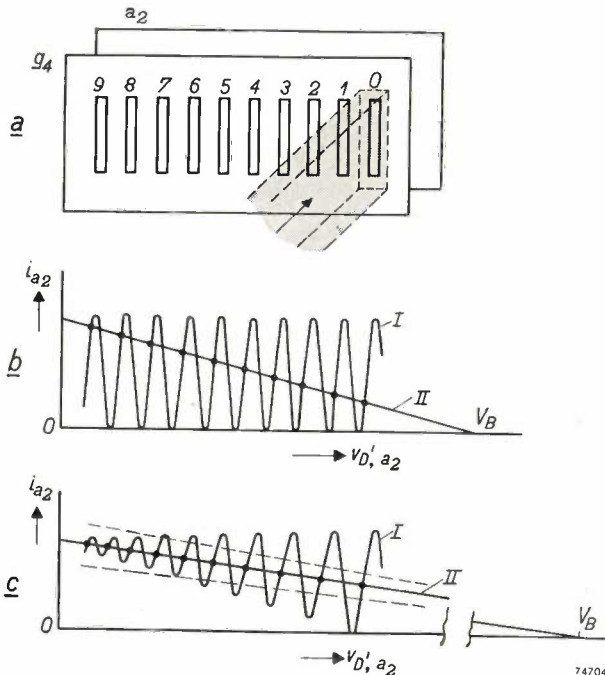


Fig. 7. (a) a_2 anode and g_1 slotted electrode of a (primitive) decade counter tube, both drawn as flat planes. The ten slots have been drawn as being of the same size and equidistant. (b) Curve I: $i_{a2} = f(v_{D',a_2})$ characteristic which might be expected in the case (a). II again represents the resistance line. (c) Due to the defocusing of the beam as it is deflected to the left, the $i_{a2} = f(v_{D',a_2})$ characteristic actually assumes the form represented by I. At small variations of R_{a2} (corresponding to the broken lines) some of the points of intersection at the left are lost. V_B must, moreover, largely exceed the value of v_{D',a_2} corresponding to the point of intersection O.

passing through the slots with low numbers, i.e. the peaks of i_{a2} become lower as the beam travels from the right to the left. The influence of the defocusing on the minima is even more serious because the beam soon becomes so wide when it is deflected that it is wider than the space between the slots. Consequently, as the beam moves towards the left, the current minima become less pronounced, i.e. the amplitude of curve I decreases towards the left.

In order to obtain the much more favourable characteristic I of fig. 6, the slotted electrode has been provided with additional apertures, which are also scanned by the beam, so that an additional current is passed which increases as the beam proceeds further to the left.

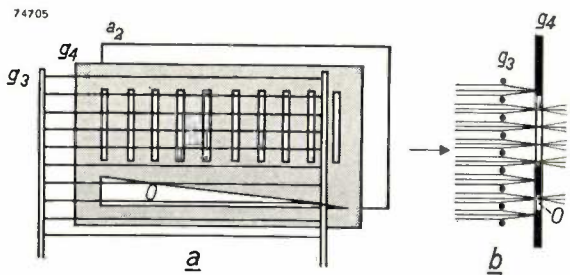


Fig. 8. (a) In order to obtain the characteristic I of fig. 6, the electrode g_3 might be provided with an additional triangular aperture O through which an additional current passes which increases as the beam is deflected further to the left. Due to the action of the suppressor grid g_3 (see the cross-section b), the magnitude of the additional current, however, depends too much on the location of g_3 with respect to the aperture O.

The most obvious solution is to make a triangular aperture in the slotted electrode (O in fig. 8). It would nevertheless be very difficult to obtain the desired improvement in this way. This is due to the presence of a suppressor grid (g_3) in front of the slotted electrode. The horizontal wires which constitute this grid give rise to a vertical variation of the current density of the beam with minima behind the grid wires and maxima behind the spaces of the wires (see fig. 8b). The "ribbon" might thus be imagined to be sliced into a number of threads situated above each other. The positions of the grid wires with respect to the narrow triangular aperture largely determine the way in which the current flowing through this aperture varies as a function of the deflection angle. This leads to difficulties in manufacture.

For this reason the idea of a narrow triangular aperture was abandoned in favour of rectangular apertures whose positions with respect to the grid wires are less critical. As shown in fig. 9, part of the beam passes through an aperture O_1 from slot 5 onwards. In positions 8 and 9 a second aperture moreover becomes operative (O_2).

A second measure applied to obtain a rising characteristic consists in making each slot slightly larger than its predecessor: starting from 2 to 9 the width gradually increases, whilst 1 is slightly lower than 2...9. Slot 0, which is made particularly wide, forms an exception, the reason of which will be shown later. To prevent the beam from bending around the edge of this wide opening — which might result in the figure 1 being indicated — slot 0 is covered with gauze.

By these measures the characteristic assumes the form as depicted ³⁾ in fig. 10. Due to the rising characteristic, V_B need not be much higher than the value of v_{D',a_2} at which the beam occupies position 0. A value of $V_B = 300$ V suffices; as will be seen from a circuit diagram to be discussed later, all other direct voltages the tube requires can be derived from the same source. The value of R_{a_2} should be approximately equal to the average slope of the characteristic, which amounts to about 1 M Ω ,

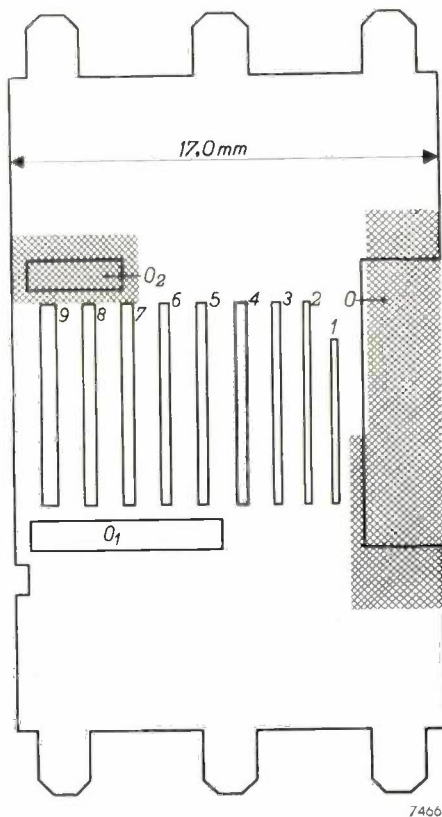


Fig. 9. Actual form of the apertures in the slotted electrode g_1 (drawn as a flat plane). The slots become larger and their distances from centre to centre increase in the sequence from 1 to 9. An additional current passes through the two apertures O_1 and O_2 of which the upper one, like slot 0, is covered with gauze.

³⁾ Plotted with an apparatus as described in Philips tech. Rev. 12, 283-292, 1950/51.

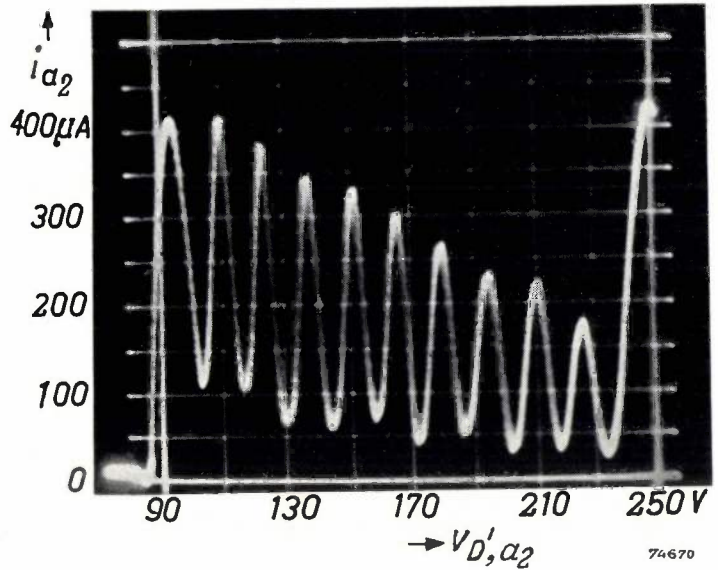


Fig. 10. Oscillographically ³⁾ plotted $i_{a_2} = f(v_{D',a_2})$ characteristic of the E1T counter tube. Cf. curve I shown in fig. 6. The oscillogram displays one peak more than the theoretical curve shown in fig. 6. This is due to the fact that in practice a peak — the one the extreme left — also occurs when the beam passes between the slotted electrode and the reset anode. This peak is of no consequence for the operation of the tube.

but owing to the large amplitude of the waves there is a reasonable play in the magnitude of R_{a_2} , although it remains desirable to keep the deviation within 1% for other reasons.

Suppressor grids, anode and fluorescent screen

The previously mentioned suppressor grid g_3 (fig. 8) serves for rejecting secondary electrons which are emitted by the slotted electrode and would otherwise proceed mainly towards the deflection electrodes. A small proportion of these electrons, however, still passes through the suppressor grid, which is particularly troublesome when the beam is in position 0 or 1. For this reason an auxiliary anode has been incorporated at the right-hand side (e in fig. 3a); a voltage of 300 V is applied to this auxiliary anode, so that it captures these electrons. The screen s, which is mounted behind and connected to the cathode, prevents primary electrons from impinging on the envelope.

A second suppressor grid (g_5 , fig. 3) is mounted between the slotted electrode and the anode and decelerates secondary electrons originating from the anode. It moreover "reassembles" the beam of primary electrons, which was sliced into threads by the first suppressor grid (fig. 8b), so that the ribbon-shaped beam again becomes homogeneous once it has passed the second suppressor grid ⁴⁾ (fig. 11). If this were not the case, the electrons which

⁴⁾ J. L. H. Jonker and B. D. H. Tellegen, Philips Res. Rep. 1, 12-19, 1945.

pass through one of the ten slots of the anode would no longer give rise to a uniformly luminescing spot, but would produce a less distinct, "crumbly" spot.

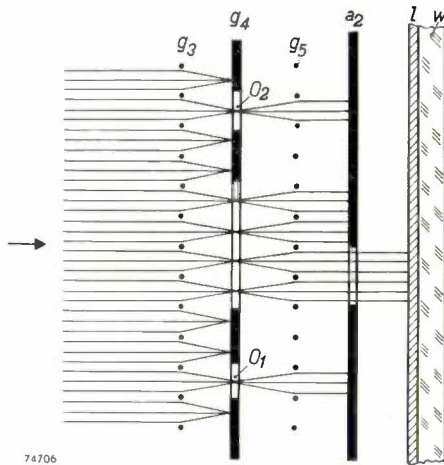


Fig. 11. Cross-section through the first suppressor grid g_3 , the slotted electrode g_4 , the second suppressor grid g_5 , the anode a_2 , the fluorescent layer l and the glass envelope w . By giving g_5 the correct pitch and position, a homogeneous beam is restored behind g_5 .

By giving the grid g_5 the same pitch as g_3 and positioning it so that its wires are situated in the same horizontal planes as those of g_3 , the beam will once again be homogeneous (fig. 11).

The fluorescing layer consists of a specially prepared type of zinc oxide with blue fluorescence and a very long life. It is applied to a conducting layer, which, during operation, is at a potential of 300 V.

In order to facilitate reading of the figures, they are placed in two rows above each other: 0...8 and 1...9 (fig. 1). The apertures in the anode are placed at corresponding positions; see fig. 12, in which the various parts of the tube are shown.

Mechanism of the displacement of the beam from 0 to 9

The positive-going voltage pulses to be counted are applied to the left deflection electrode D via a blocking capacitor, as shown in fig. 5.

In order to understand how the beam is shifted to a following position at each pulse, it should be recognized that the characteristic I shown in fig. 6 is applicable to a constant voltage v_D at the left deflection electrode, and that the angle of deflection is a function of $v_D - v_D$. An increase of v_D by an amount V_i therefore corresponds to the line I being shifted to the right over a distance corresponding to V_i (fig. 13).

As a starting point it is assumed that the beam occupies one of its stable positions, for example

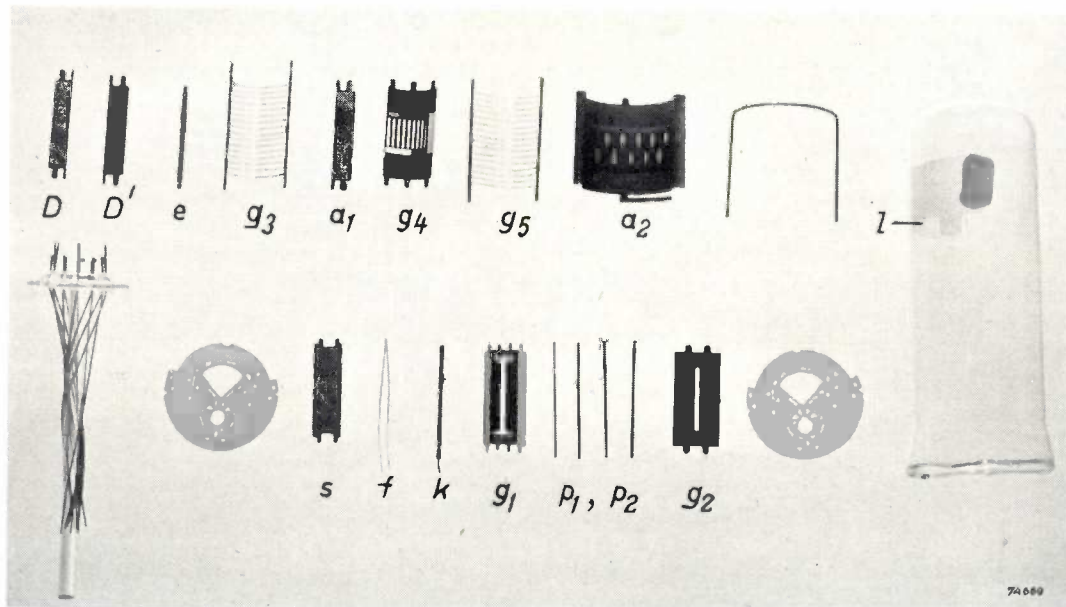


Fig. 12. Component parts of the EIT counter tube. The references have the same meaning as in fig. 3.

Lower row from left to right: glass base with pumping stem, mica support, screen (s), heater (f), cathode (k), control grid (g_1), four focusing electrodes (p_1, p_2), accelerating electrode (g_2), second mica support.

Upper row, from left to right: the two deflection electrodes (D, D'), auxiliary anode (e), first suppressor grid (g_3), reset anode (a_1), slotted electrode (g_4), second suppressor grid (g_5), anode (a_2), resilient bracket (to establish contact with the conductive layer on the envelope), envelope with fluorescent layer (l).

position 0. If a positive pulse is now applied to the left deflection electrode, so that v_D is temporarily increased, the beam will tend to move to the left (fig. 3a); the number of electrons passing through

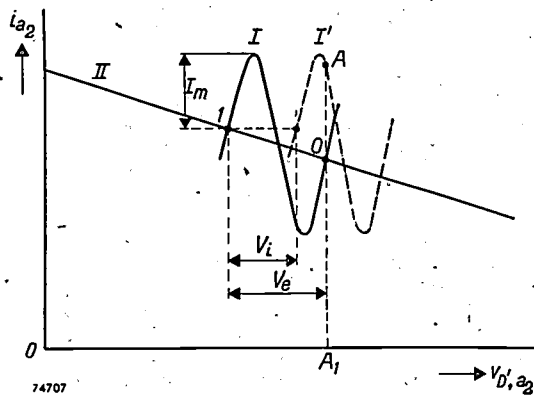


Fig. 13. $i_{a_2} = f(v_{D,a_2})$ characteristic I of the counter tube for a given voltage v_D on the left deflection electrode; I' the same characteristic for a voltage $v_D + V_i$ on the left deflection electrode. II resistance line. I_m amplitude of the waves. V_e horizontal distance between two adjacent stable points of intersection.

slot 0 then decreases, i.e. the anode current i_{a_2} is reduced. If no stray capacitances were present, the decrease of i_{a_2} would result in a rise of the potential of the anode and of the right deflection electrode connected to it, and this increase of v_{D',a_2} would counteract the deflection to the left, so that the beam would be retained at position 0.

In practice, however, the stray capacitance to earth of the electrodes a_2 and D' and their wiring, represented by C_{a_2} in fig. 5, is shunted across R_{a_2} and impedes sudden changes of the potential of a_2 and D' . Provided the condition is satisfied that the leading edge of the pulse is sufficiently steep, the potential v_{D',a_2} may be considered constant during the rise time ϑ_1 of the pulse (fig. 14). This amounts to the line I of fig. 13 being shifted to the right (I') over a distance equal to the amplitude V_i of the pulse, the anode current thus assuming the value A_1A . This differs from the original value, whereas the current flowing through R_{a_2} still has its original value; the difference is supplied by the capacitance C_{a_2} .

Provided the second condition is also satisfied, namely that the decay time ϑ_2 , during which the pulse decays from V_i to zero, is sufficiently long, the characteristic I' will gradually return to I , and A will be shifted to the adjacent stable point of intersection between I and II , i.e. point I . In order to ensure that the beam is shifted just one step to the left, a third condition must be imposed to the beam, namely that the amplitude V_i should be roughly equal to the voltage difference

V_e which corresponds to the horizontal distance between two adjacent stable points of intersection and amounts to approximately 14 V; it will be clear that at too small an amplitude the beam will return to its original position, whereas at too large an amplitude it will be advanced two or more steps.

As a result of the deflection being asymmetrical, the average potential in the space between the deflection electrodes drops each time the beam is shifted a step further to the left, since v_{D',a_2} decreases and v_D remains constant. As a consequence the deflection sensitivity increases, i.e. the reduction of v_D corresponding to a given increase of the deflection is smaller at the left than at the right. In the case of the primitive slotted electrode shown in fig. 7a, with equidistant slots, this means that the distances between the points of intersection of I and II become smaller in the direction from the right to the left (see fig. 7c), hence that V_e decreases from the right to the left. This is undesirable as it decreases the tolerance in the amplitude V_i of the pulses. This objection has been met with by increasing the distances between the centres of the slots in the direction from 0 to 9 in such a way that the gradually increasing deflection sensitivity is exactly compensated.

In practice the conditions imposed on the pulse may be formulated as follows:

$$\text{Interval } \vartheta_1: \quad \frac{dv_i}{dt} \geq 3 \frac{I_m}{C_{a_2}}, \dots \dots \dots (2)$$

$$\text{Interval } \vartheta_2: \quad -\frac{dv_i}{dt} \leq 0.3 \frac{I_m}{C_{a_2}}, \dots \dots \dots (3)$$

where v_i denotes the instantaneous value of the pulse amplitude and t the time. Theoretically, it might be expected that the amplitude of V_i may assume a value between $\frac{1}{2}V_e$ and $\frac{3}{2}V_e$, but the margin becomes smaller due to several causes. In fact, the following condition is valid:

$$\left. \begin{aligned} 11.5 \text{ V} < V_i < 16 \text{ V} \\ \text{or } V_i = 13.6 \text{ V} \pm 18\% \end{aligned} \right\} \dots \dots \dots (4)$$

When it is assumed that $I_m = 0.1 \text{ mA}$ and $C_{a_2} = 15 \text{ pF}$, it can be calculated that v_i must

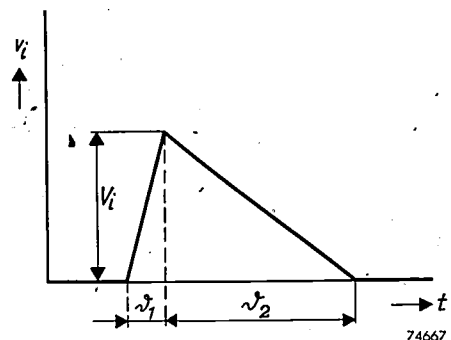


Fig. 14. Pulse voltage v_i as a function of the time t . V_i denotes the amplitude, ϑ_1 the rise time, ϑ_2 the decay time.

increase not slower than $20 \text{ V}/\mu\text{sec}$ and subsequently decrease not faster than $2 \text{ V}/\mu\text{sec}$. At these extreme values of the slopes and $V_i = 14 \text{ V}$, $\theta_1 = 0.7 \mu\text{sec}$ and $\theta_2 = 7 \mu\text{sec}$.

If the counting pulses do not satisfy the imposed conditions they must be applied to a pulse shaper, i.e. a device which produces a pulse of the required form and amplitude for each pulse applied. Such a device will be referred to later.

Circuit for counting rates up to 30,000 per second

As previously shown, each tenth pulse applied to a counter tube should have two consequences: firstly the beam in the tube considered should be reset from position 9 to position 0, and secondly a pulse should be fed to the following counter tube to advance its beam one step. Both events are

purpose. On closer investigation some of these circuits appeared to have the disadvantage of being suitable only for fairly low counting rates, whilst others produce faulty counts when the resistors or capacitors deviate very slightly from the rated values or when a counter tube is replaced.

These objections are not applicable to the carefully investigated counter circuit described below; a practical example of a decade counter built on these lines is shown in fig. 2. This circuit is characterized by the fact that each counter tube is followed by a pulse shaper, which feeds a resetting pulse to the preceding counter tube and a counting pulse to the following tube. The first counter tube is, moreover, preceded by a pulse shaper (of a slightly different type) which is commonly used in counter circuits and serves for

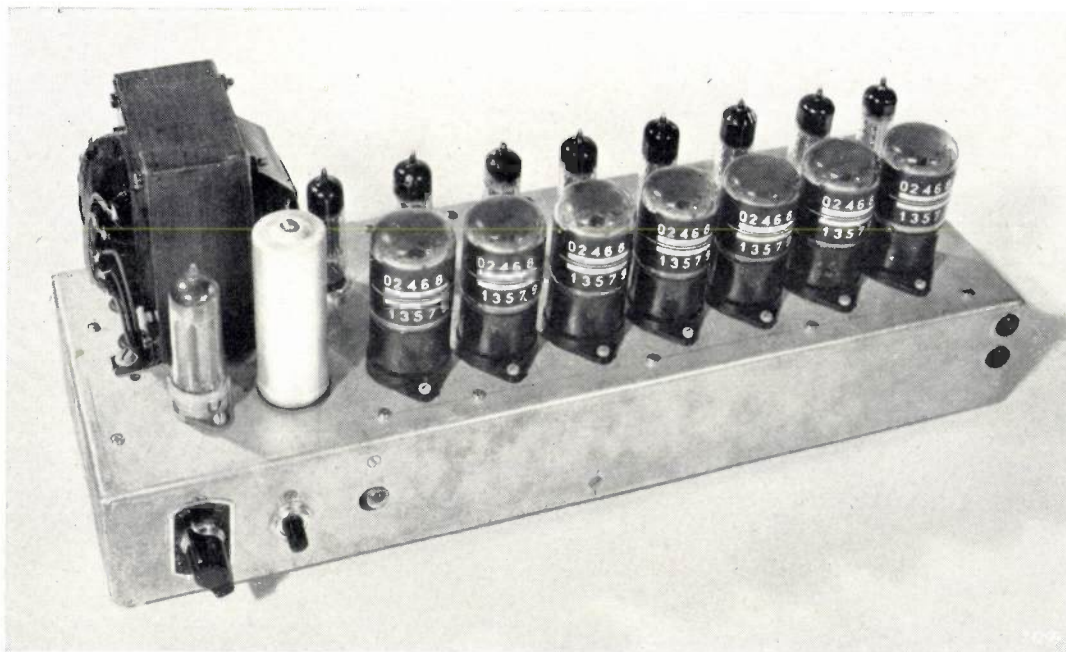


Fig. 15. The decade counter depicted in fig. 2 with the cover removed. Behind the seven EIT counter tubes the eight double triodes E90CC, which act as pulse shapers, can be seen. At the left is the rectifier valve AZ 41 which provides the direct current for the entire apparatus (300 V, approximately 50 mA).

initiated by the fact that every tenth pulse initially deflects the beam occupying position 9 still further to the left, so that it leaves the slotted electrode and impinges on the reset anode.

The reset anode is connected to the $+300 \text{ V}$ line via a resistor, so that its potential (v_{a1}) depends on i_{a1} , i.e. on the extent to which it is struck by the beam. The variations of v_{a1} are now used for initiating the resetting and for ensuring that a pulse is passed to the following counter tube.

Several circuits have been designed for this

converting the initial pulses into pulses of the required form and amplitude. Both types of pulse shaper incorporate a double triode (E90CC), so that for n counter tubes ($n+1$) E90CC tubes are required (see fig. 15).

The two types of pulse shaper will now be discussed separately.

Interstage pulse shaper

The pulse shaper which follows each counter tube is triggered each time it receives a pulse from

the reset anode of the preceding counter tube. It then feeds a counting pulse of the required form and amplitude to the following counter tube and, moreover, ensures that the current in the preceding tube is temporarily suppressed during a sufficient length of time to allow the anode voltage of this tube to rise to the value corresponding to position 0, after which the beam is restored.

Further details follow from the circuit shown in

potential rise of g' , current starts to flow through T' , and since the anode resistor of T' has a much smaller value than that of T ($R_{a'} = 3300 \Omega$, $R_a = 39,000 \Omega$), this current largely exceeds that which flowed through T . As a result of this larger current, the cathode potential rises to such an extent that T is completely cut off.

Meanwhile the coupling capacitor C_2 is charged. The charging current flows through the resistor R_4 , thus producing a voltage drop which acts as a positive grid bias for T' . The charging current decreases exponentially, and so does the potential of g' . The cathode potential follows that of g' and

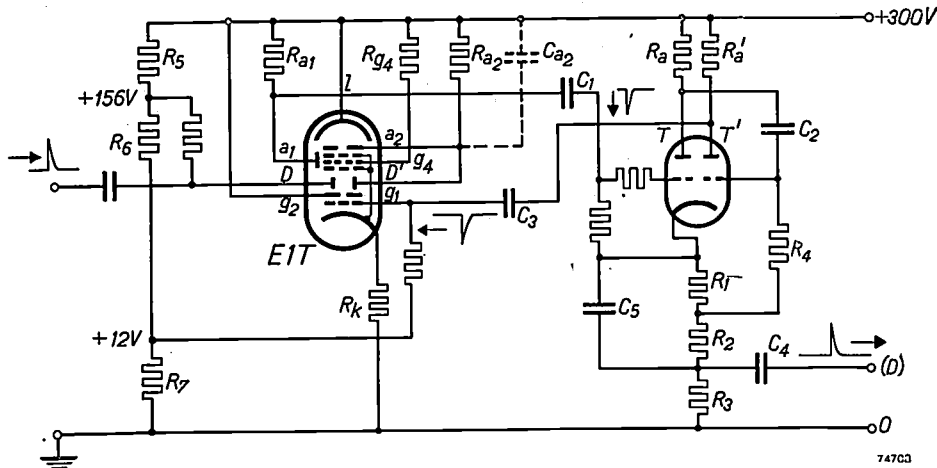


Fig. 16. Circuit of a counter tube and the following pulse shaper. The latter is a monostable (one-shot) multivibrator (flip-flop) incorporating a double triode $T-T'$ (type E 90 CC).

At each tenth pulse applied to D the beam is advanced to a_1 , which results in a negative pulse being fed to the grid of T via C_1 . The multivibrator is thus triggered and brought from the stable condition (T conducting, T' cut off) into the quasi-stable condition (T' conducting, T cut off), but after a short interval it returns to the stable condition. During this operation two pulses are supplied: across R_3 a positive pulse is produced, which is fed to the following counter tube, whilst at the anode of T' a negative pulse is produced which is fed to the control grid (g_1) of the preceding counter tube; its beam is thereby temporarily suppressed and occupies position 0 after being restored.

R_a is the load resistor of T (39,000 Ω), $R_{a'}$ is the load resistor of T' (3300 Ω).

fig. 16, in which one stage (counter tube and accessory double triode) is represented. The E90CC, in combination with a number of resistors and capacitors, forms a monostable ("one-shot") multivibrator, i.e. a multivibrator with one stable and one quasi-stable condition. The stable condition is that in which only triode section T passes current. When a negative voltage pulse is applied to the grid of T , this section is cut off and the other section, T' , becomes conducting. From this quasi-stable condition the multivibrator automatically returns to the initial condition.

The operation of the monostable multivibrator may be explained as follows. The current which flows through T produces in the part R_1 of the cathode resistor ($R_1 + R_2 + R_3$), a voltage drop which acts as negative grid bias for T' and ensures that this section is cut off. The anodes and the grids of T and T' will be denoted by a and a' , and g and g' respectively. When the current flowing through T is decreased by a negative pulse on g , the potential of a and thus also that of g' rises, since g' is coupled to a via a capacitor (C_2). Due to the

after some time — depending on $(R_a + R_4) C_2$ — reaches the value at which T again becomes conducting. The potential of a then drops and so does that of g' . Section T' is cut off and C_2 is discharged via T , the stable condition thus being re-established.

The negative pulse which triggers the multivibrator is derived from the reset anode via a coupling capacitor (C_1). The potential of this anode obviously drops when it is struck by the beam.

The sudden voltage drop which occurs at the anode of T' when current starts to flow through T' , is used for temporarily suppressing the beam in the counter tube (in this way also terminating the negative pulse by which the multivibrator was triggered). This is achieved by coupling the anode of T' via a capacitor (C_3) to the control grid (g_1) of the counter tube. This grid has normally a voltage of +12 V with respect to earth, derived from a voltage divider ($R_5 - R_6 - R_7$) between the +300 V line and 0. By reducing the voltage on g_1 to -15 V or

an even lower value, the beam is suppressed.

While the beam still occupies position 9 (anode current = I_{a9}), the anode potential of the counter tube is $V_{a9} = V_B - I_{a9}R_{a2}$. Now V_{a9} is the lowest of the various potentials $V_{a0} \dots V_{a9}$ that the anode assumes in the positions 0...9 — as shown by the oscillograms of *figs 17a* and *b* — since I_{a9} is the largest anode current. As the resistor R_{a2} , shunted by the stray capacitance C_{a2} , is connected between the anode and the +300 V line, the anode potential $v_{D,a2}$ rises exponentially as soon as the anode current has become zero (as a result of the beam having been advanced to the reset anode and subsequently having been suppressed). The beam must be suppressed during such a length of time that $v_{D,a2}$ has sufficient opportunity to rise to a value which exceeds $V_{a0} = V_B - I_{a0}R_{a2}$, where V_{a0} and I_{a0} denote the anode voltage and anode current at position 0 of the beam. When the beam is now restored after the negative pulse on the con-

trol grid has ceased, the anode and the right deflection electrode have reached a potential which exceeds the potential corresponding to position 0. As a consequence the beam is deflected to the extreme right and from this position it automatically travels to the adjacent stable position, i.e. position 0. The resetting is thus completed.

The same principle is used for resetting all counter tubes to zero before counting starts. By means of a press-button switch (*fig. 2*) a negative voltage is applied to all control grids g_1 , as a result of which the beams are suppressed and occupy position 0 when they are restored. The circuit required for this purpose has been omitted in *fig. 16*.

Finally, the pulse which should be fed to the following counter tube will be dealt with. For this purpose the voltage drop is used which is produced across the part R_3 of the cathode resistor of $T-T'$ when, owing to the multivibrator being triggered, a larger current starts to flow through T' than previously flowed through T . The pulse is applied to the left deflection electrode of the following counter tube via a coupling capacitor (C_4).

The slope of the leading edge of this pulse depends on the speed with which the multivibrator changes from the stable to the quasi-stable condition. If the capacitor C_5 were absent, the strong negative feedback caused by the presence of the cathode resistor ($R_1 + R_2 + R_3$) would delay the triggering of the multivibrator to such an extent that the pulse would not be steep enough. Large negative feedback is, however, necessary to ensure good stability (insensitivity to changes of the valve characteristic, etc.). To solve this difficulty, $R_1 + R_2$ is by-passed by the capacitor C_5 , the capacitance of which is so chosen that only the cathode resistor R_3 is operative for "rapid" changes, whereas the complete resistor ($R_1 + R_2 + R_3$) is operative for slow changes.

The cathode resistor (R_k) of the counter tube also serves for increasing the stability.

The circuit of *fig. 16* occurs seven times in the decade counter depicted in *fig. 15*. This circuit may be built as a separate unit which can be plugged into a tube holder. In this way counters can easily be assembled. *Fig. 18* shows an example of such a unit ⁵⁾.

Amplitude and duration of the pulses; counting rate

The positive pulse supplied by the pulse shaper must satisfy the conditions (2), (3) and (4) formulated above: sufficiently high slope of the leading edge, sufficiently low slope of the trailing edge, and an amplitude of $13.6 \text{ V} \pm 18\%$. The negative pulse, see *fig. 19*, must be sufficiently large to suppress the beam during at least the time τ that is required by $v_{D,a2}$ to rise from the value V_{a9} to

⁵⁾ This unit is not in production.

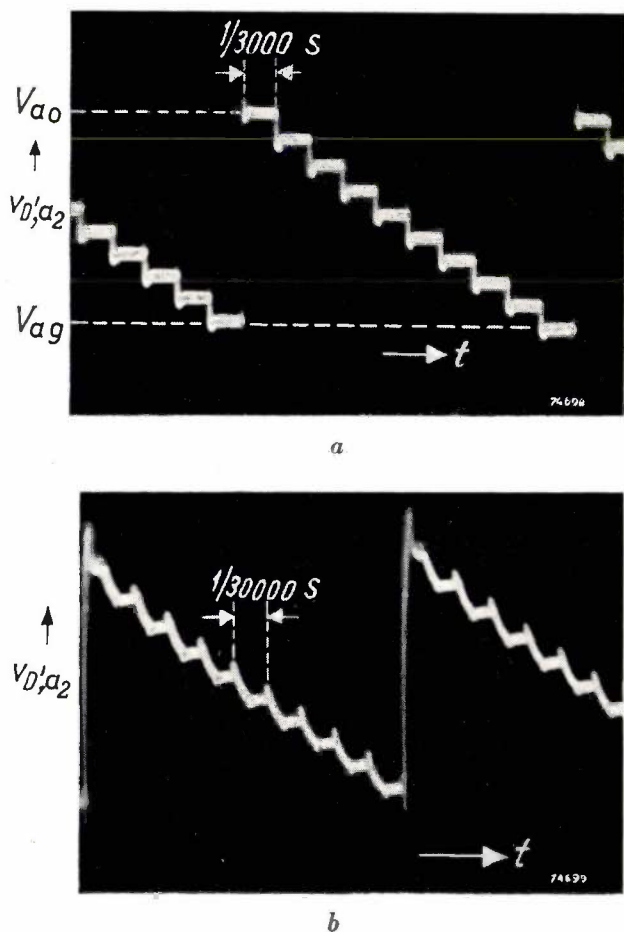


Fig. 17. Oscillograms of the anode voltage $v_{D,a2}$ of the counter tube shown in *fig. 16*, as a function of the time t . (*a*) has been registered at 3000 and (*b*) at 30,000 pulses per second. In position 0: $v_{D,a2} = V_{a0} = 240 \text{ V}$; in position 9: $v_{D,a2} = V_{a9} = 95 \text{ V}$. Between 9 and 0 the beam is suppressed and $v_{D,a2}$ rises exponentially.

V_{a0} , but its duration should, on the other hand, not exceed τ considerably, since this would affect the counting rate.

The time τ is given by the following equation:

$$V_B - V_{a0} = (V_B - V_{a9}) e^{-\tau/R_{a2} C_{a2}}$$

By substituting $V_B = 300$ V, $V_{a0} = 240$ V, $V_{a9} = 95$ V, $R_{a2} = 1.05$ M Ω and $C_{a2} = 16$ pF in this equation, τ is found to be 21.3 μ sec. Making allowance for a safety margin, the condition is imposed that the pulse must not be shorter than 23 μ sec. Taking the same tolerance ($\pm 18\%$) for the pulse duration as for the amplitude, an average duration of 27.2 μ sec is obtained, the extreme values being 23 and 32 μ sec. Allowing another 1.3 μ sec as a margin, the longest pulse duration corresponds to a maximum repetition frequency of $10^6/(32 + 1.3) = 30,000$ pulses per second.

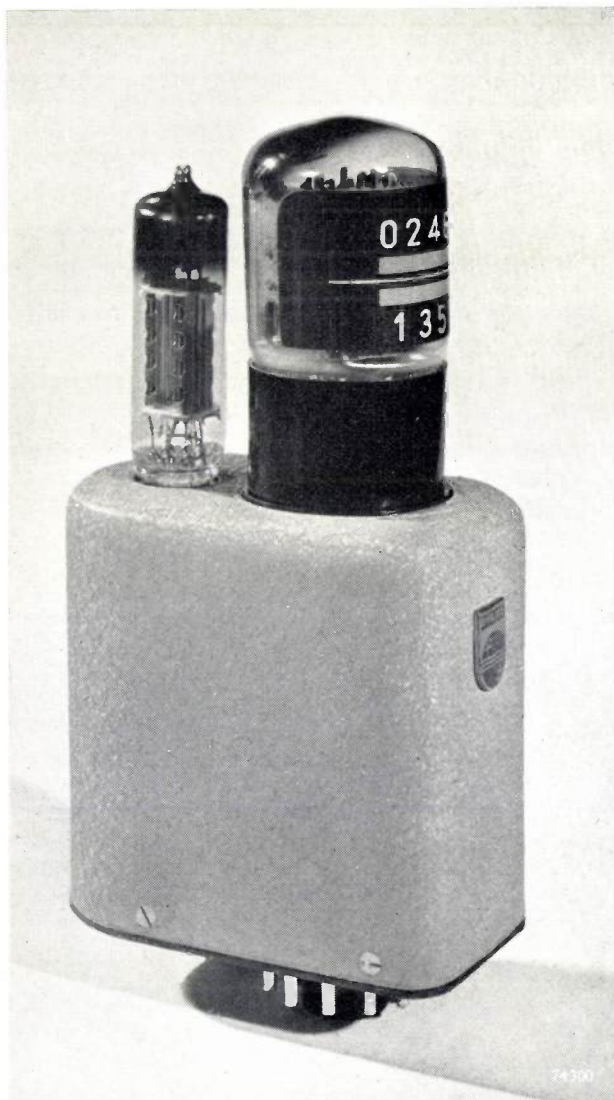


Fig. 18. A decade counter tube with its accessory pulse shaper connected according to the circuit of fig. 16, in the form of a unit⁵⁾ to be inserted in a tube holder.

The conditions regarding the form, amplitude and duration of the pulse can be met by a suitable choice of the component values. The question remains, however, whether these conditions are still satisfied when these values are subject to slight changes. In the circuit of fig. 16, by way of experiment, all resistances were changed simultan-

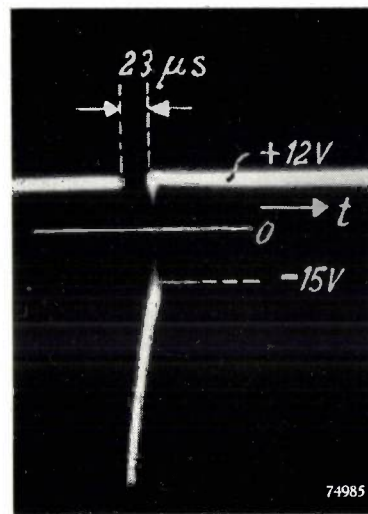


Fig. 19. Oscillogram of the negative pulse with which the pulse shaper (fig. 16) temporarily suppresses the beam of the corresponding counter tube.

ously by 2% and all capacitances by 5%; the sign of each change was such that the influence on the final result was a maximum. The change of the pulse duration proved to be 7.5% and that of the amplitude appeared to be 8%; both changes thus remain well within the tolerance of 18%.

Input pulse shaper

As the original pulses which are to be counted do not, as a rule, satisfy the imposed conditions, the first counter tube must be preceded by a pulse shaper. It might be expected that the previously discussed multivibrator can be used for this purpose, but a closer investigation reveals that it is necessary to extend this circuit.

The interstage pulse shapers are triggered by negative pulses which are derived from the preceding counter tube (fig. 16); the duration of these pulses is but a fraction of 1 μ sec, that is, much shorter than the duration of the output pulses (23-32 μ sec). The pulse shaper at the input, on the other hand, is triggered by the original pulses, the duration of which may be considerably longer than 23 μ sec, particularly so at low counting rates. For reliable operation a multivibrator requires triggering pulses of shorter duration than the pulses delivered.

This difficulty has been overcome by differentiating the original pulses by means of a capacitor C_d and a resistor R_d (fig. 20). If the form of the original pulse is, for example, as drawn in fig. 21a, a voltage will be produced across R_d , which — provided $R_d C_d$ has been chosen small enough — approximates the form depicted in fig. 21b, remaining negative during a shorter length of time than the

remain within certain tolerances, a counting rate of 30,000 pulses per second can be obtained with certainty.

All E1T tubes are tested at this counting rate; during these tests the pulse amplitude and the anode resistor are varied in order to guarantee reliable operation of the tubes in the circuit discussed at 30,000 pulses per second. This counting

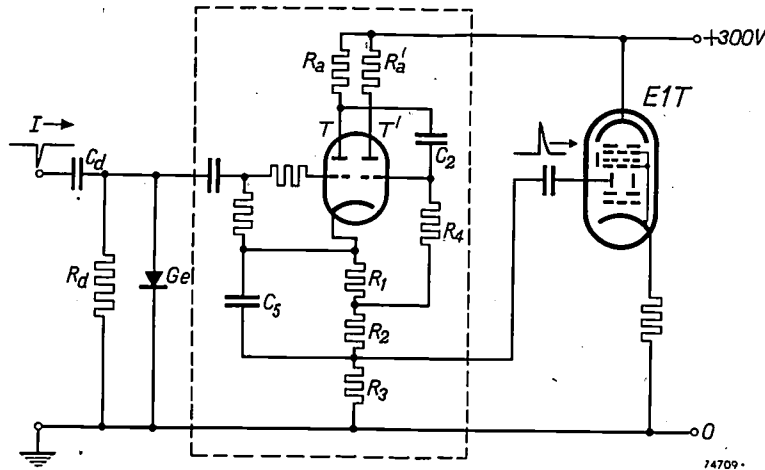


Fig. 20. Circuit of the input pulse shaper. I applied negative-going counting pulses. C_d - R_d differentiating network. Ge germanium diode (type OA 55), which prevents a positive voltage occurring across R_d . The circuit within the rectangle corresponds to that shown in fig. 16 of the other pulse shapers. At the right of this circuit is the first counter tube.

original pulse. The negative voltage is now, however, followed by a positive voltage which suddenly drops to zero at $t = t_2$. This sudden variation might trigger the pulse shaper again and give rise to a faulty count. This is why R_d is shunted by a germanium diode (fig. 20), which prevents the voltage across R_d from becoming positive.

The input pulse shaper need supply pulses for operating the first counter tube only. The condition that the pulse duration must be greater than τ is not applicable to these pulses, so that it was possible to reduce the duration of these output pulses to 13.5 μ sec.

It would lead too far to discuss the other (small) differences between this pulse shaper and the others, and it will suffice to quote the conditions which must be imposed to the original pulses (assuming these to have roughly a square wave-form):

- Amplitude 20 V to 50 V,
- Leading-edge duration < 13.5 μ sec,
- Interval between end of one pulse and beginning of next $\geq 10 \mu$ sec.

These conditions can as a rule easily be satisfied, and provided the values of the circuit components

rate is however by no means the highest rate obtainable with the tube. In the laboratory it has proved possible to obtain counting rates of the order of 100,000 pulses per second and, by using a triode-hexode and a secondary-emission valve as

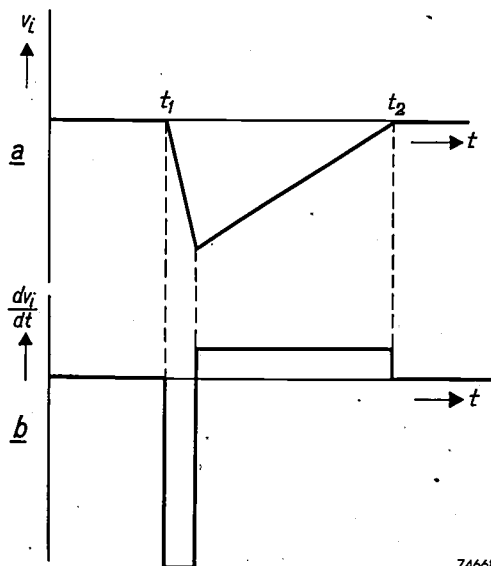


Fig. 21. Form of the original pulses (a) and of the pulses (b) after having been differentiated by C_d - R_d (fig. 20). The positive part in (b) is clipped by the germanium diode.

auxiliary valves, counting rates of millions of pulses per second have even been obtained ⁶⁾. For achieving these record performances the circuit parameters were adjusted to suit the individual tubes. On the other hand, the standard circuit discussed above will count up to 30,000 pulses per second without any special adjustment being made to allow for the unavoidable differences between production tubes.

⁶⁾ See the second article quoted in footnote ²⁾.

Summary. A decade counter tube (EIT) has been designed which gives visual indication of the counts on the tube itself. The EIT is a cathode-ray tube — roughly the size of a receiving valve — whose electron gun produces a ribbon-shaped electron beam with a current of approximately 1 mA. A supply voltage of 300 V suffices for this tube. The beam impinges on an electrode with ten slots, behind which there is an anode with ten apertures. When the beam passes through a slot and an aperture, it impinges on a fluorescent layer with which the envelope

is lined; at this spot a luminescent mark is produced adjacent to the corresponding figure (0.....9).

On its way from the gun to the slotted electrode the beam passes between two deflection electrodes. The slots and some auxiliary apertures in the slotted electrode are so dimensioned that the anode current as a function of the deflection voltage assumes a number of maxima and minima which increase as the beam is further deflected from 0 to 9. By feeding the anode and the one deflection electrode (that nearest to the figure 0) via a common resistor of suitable value, the beam is given ten stable positions of equilibrium, corresponding to the figures 0.....9. The beam is advanced to the following position by applying a positive pulse to the other deflection electrode.

At each tenth pulse the beam impinges on the reset anode. A suitable circuit ensures that the resulting voltage surge at this anode rapidly resets the beam to position 0, and, moreover, that a pulse is fed to the following counter tube, thus advancing its beam one step. The second tube thus counts the decades, and proceeding in this way, a third tube can be made to count the hundreds, a fourth tube the thousands, etc.

The slope and amplitude conditions which the pulses must satisfy are discussed. A circuit is outlined by means of which 30,000 pulses per second can be counted. In this circuit pulse shapers are used; the input pulse shaper converts the initiating pulses into pulses of the required form and amplitude, whilst the interstage pulse shapers initiate the resetting and feed a pulse to the following counter tube. All EIT counter tubes are tested at 30,000 pulses per second.

NEW DEVELOPMENTS IN THE IMAGE ICONOSCOPE

by J. C. FRANCKEN and H. BRUINING.

621.385.832:621.397.611

Although a description of the image iconoscope was given in this Review not long ago, it is now possible to publish details of two important improvements. One of these consists of the means of avoiding what is known as an ion spot or ion burn and, simultaneously, of reducing to a negligible degree a certain distortion in the electron-optical image. The other improvement refers to a new design of focusing coil which allows of continuous electrical control of the field of view without moving the camera. (This can be done optically only with the aid of very complicated lens systems.)

In Europe the image iconoscope is the most generally used camera tube for television. As explained in a previous article¹⁾, this tube performs its task — the conversion of the optical image or scene into an electrical signal — in three stages.

First the optical image is converted into a photo-emission image by means of the photo-cathode; in the second stage this image is reproduced in the target by an electron lens and, owing to the effect of secondary emission, a potential pattern is created on the target. Finally, a scanning device erases this potential pattern and simultaneously produces the video signal which is, in effect, the electrical translation of the original optical image.

The new development which is the subject of this article relates to the second stage, that is, the electron-optical projection of the photo-cathode onto the target. After mentioning an improvement in the internal design of the tube itself, which results in a much longer life and less distortion, an improvement to one of the accessories will be discussed which permits continuous variation of the electron-optical magnification.

Disadvantages of the ordinary electron lens

There are one or two disadvantages inherent in the ordinary type of electron lens, a full description of which was published in article I.

From the electron-optical aspect, the greatest defect is a certain amount of rotational distortion, although this in itself is not usually very noticeable. A more serious shortcoming originates in the form of the electrostatic field of the lens (see *fig. 1*). In the absence of a magnetic field, this electric field has the same effect as a negative lens and causes the paths of the emitted photo-electrons to diverge. Positive ions, however, which may be produced by impact between the photo-electrons and residual

gases in the tube, are accelerated in the direction of the photo-cathode and their paths converge to form a spot of about 1 mm diameter in the centre of the cathode. Actually, the same thing happens when the magnetic field is present, as this is too weak to have any influence on the paths of the ions. In time, owing to this bombardment, the photo-sensitive layer is damaged and locally becomes less sensitive. An unpleasant dark spot is then visible in the televised picture, this being known as an ion spot.

So far, this has always been the factor that determines the life of the camera tube, and its occurrence depends upon circumstances which are very difficult to predict. In some tubes the spot becomes noticeable after several hundred hours' use, but there have been instances in which a tube

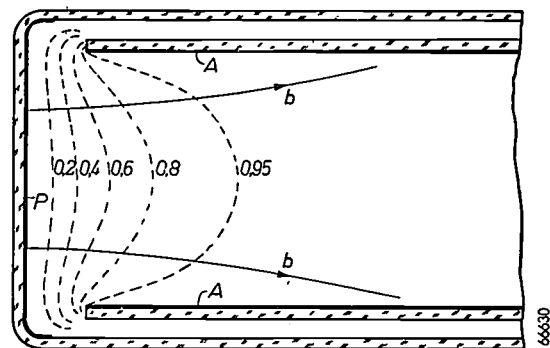


Fig. 1. Cross-section of the photo-cathode end of the old type of image iconoscope (type 5854/00). An accelerating electric field is present between the photo-cathode *P* and the anode *A*. Broken lines: equipotential planes (the potential of *P* is zero, that of *A* is taken to be unity). *b* = electron paths in the absence of other fields.

has thus become useless after only 50 hours' service. The ion spot can be particularly annoying in outside programmes, especially sports events; the camera is then constantly moving and contrasts in the scene (e.g. the sports ground) are usually quite low. The dark spot is then much more distracting than in programmes televised from the studio,

¹⁾ P. Schagen, H. Bruining and J. C. Francken, Philips tech. Rev. 13, 119-133, 1951 (No. 5), hereinafter referred to as article I.

where the camera is not moved so much and where the background (e.g. a back-drop), provides much greater contrasts.

Clearly then, there is an urgent need for obviating the occurrence of ion burn.

The use of a gauze

Both the defects mentioned can be remedied by closing the end of the anode cylinder with a very fine metal gauze (see *fig. 2*). This entirely removes

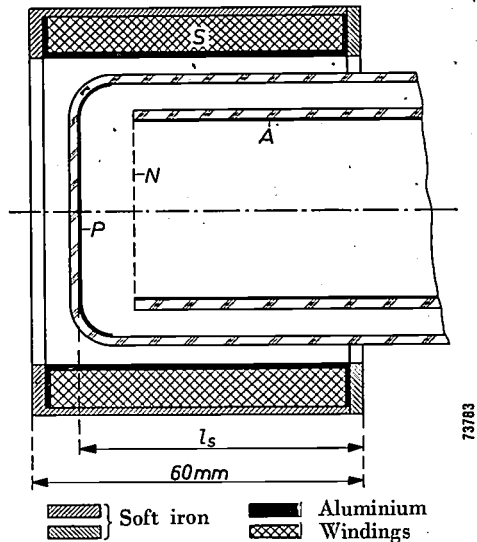


Fig. 2. Cross-section of the photo-cathode end of the improved image iconoscope (type 5854/02). P = photo-cathode. A = anode. N = fine metal gauze. S = focusing coil (diagrammatic).

the possibility of the occurrence of an ion spot. The electrostatic field between the photo-cathode and the anode is then uniform, and the positive ions are evenly distributed over the whole surface. In the long run the ion bombardment will still result in deterioration in the sensitivity, but the time involved before this happens will obviously be much longer than when bombardment is concentrated, as is the case with the field in its original form. Moreover, the image does not suffer in any way, since the decrease in sensitivity is uniformly spread over the whole area of the photo-cathode and is not manifested as a spot. Life tests have shown that there is no trace of an ion spot after several hundred hours' operation.

The question now arises what effect this modification in the electrostatic field will have on the quality of the electron-optical image (definition, magnification, distortion), as well as on the sensitivity.

Definition

On first thoughts it would appear that defini-

tion of the image would be bound to suffer, since the apertures in the gauze function as so many electron lenses. Experimentally, however, it has been found that there is no appreciable loss of definition in comparison with the old system, provided only that the mesh of the gauze is fine enough.

In the Philips image iconoscope (see article I) the limit occurs at a mesh of about 50 μ . The space between cathode and gauze is about 10 mm and the diameter of the particular cathode face is 20 mm.

Magnification

The magnification of the image differs from that obtained with a tube not fitted with the gauze. When the same coil is used for the two types of tube, the location of the photo-cathode in the magnetic field being the same in each case, the magnification is greater in the tube with the gauze (see *fig. 3*).

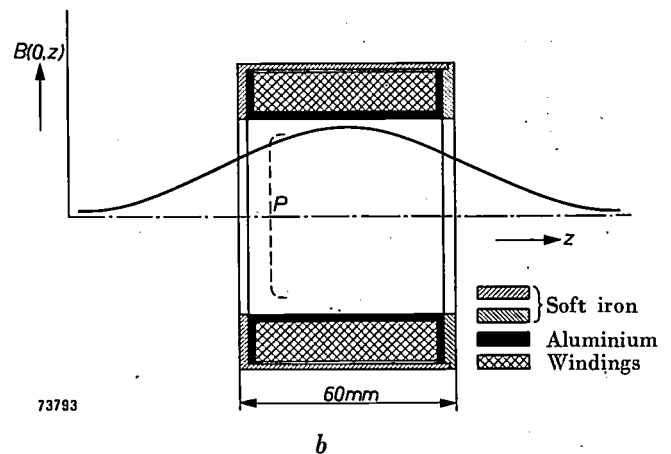
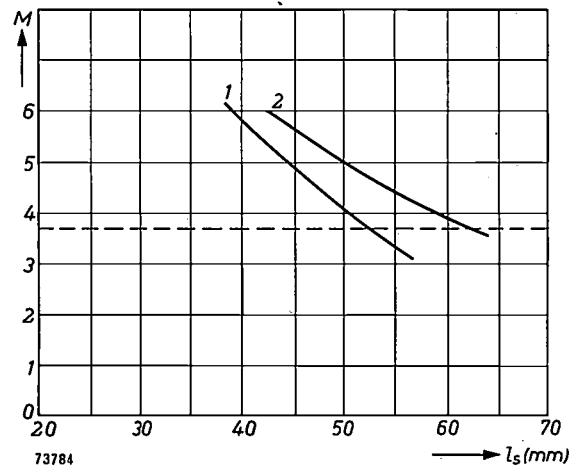


Fig. 3. a) Electron-optical magnification M of the image iconoscope type 5854 plotted against the distance l_s (see *fig. 2*). Curve 1 refers to the tube without gauze, curve 2 to the tube with gauze. The relevant focusing coil is shown in (b), in which the magnetic induction $B(0, z)$ along the axis (z) of the coil is also given. P = relative position of the photo-cathode.

In itself, this is a disadvantage, since from the point of view of sensitivity, it is important to keep the magnification in a given tube as low as possible. For a certain size of target the greatest possible area of the photo-cathode is then utilised, which means that higher sensitivity of the whole system can be achieved than if the effective area of the photo-cathode were small. (This will be explained in greater detail in a subsequent article dealing with the sensitivity of camera tubes.)

In the old type of image iconoscope (type 5854/00) an area of 12 mm \times 16 mm of the photo-cathode was used. As the dimensions of the target were 45 mm \times 60 mm, the magnification was 3.75 \times . In the tube with the gauze (type 5854/02) the same magnification can be obtained using the same coil by moving the latter nearer to the target end of the tube. As the electrons are now subjected to the magnetic field for a greater part of their travel, the magnification is decreased so that the magnifying effect of the gauze can be annulled (see fig. 3a). Under such conditions, however, considerable field curvature would result, in consequence of which the edges of the image are no longer sharp. As already pointed out in article I, this is due to the fact that the lines of magnetic force are then no longer perpendicular to the photo-cathode.

It was accordingly found necessary to design a new focusing coil for the tube containing the gauze, to give the desired magnification whilst retaining

high definition at the corners. Dimensions for this coil were determined very simply from the experiments described below.

Rotational distortion

Theoretical considerations have led to the conclusion that the major source of rotational distortion is to be found in the vicinity of the photo-cathode, and it was somewhat surprising to find that this source of distortion originates mainly in a lack of uniformity of the electric field ²⁾.

In principle it should be possible to correct this defect by making the magnetic field non-uniform as well, viz. by making the magnetic field strength increase towards the edges. This idea has been worked out elsewhere ³⁾.

We have pursued another course, however, i.e. that of making the electric field uniform (by means of the gauze), this being at the same time, as already mentioned, an effective means of eliminating the ion spot.

The results, as far as distortion is concerned, obtained from a tube containing the gauze and fitted with the new coil, are demonstrated by the photographs shown in fig. 4. Fig. 4a was obtained from a tube without gauze, using the old type of coil.

²⁾ J. C. Francken, *Electron Optics of the Image Iconoscope*, Thesis, Delft, 1953.

³⁾ German Patent Applic. Class 21g 29/40 No. F3126 and No. P 5137, published 28th February 1952 and 30th April 1952, respectively.

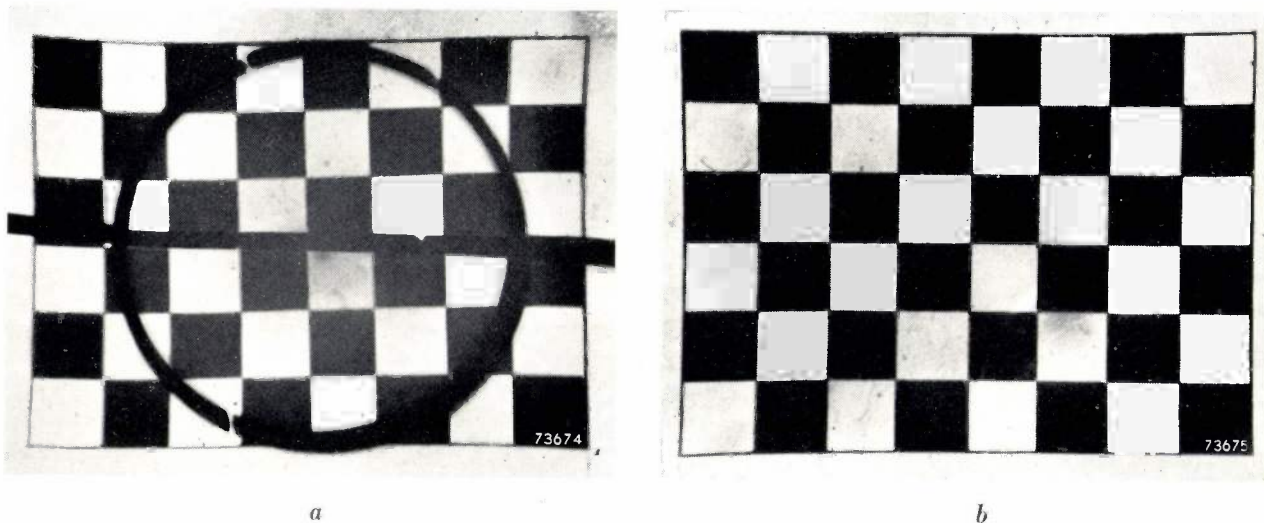


Fig. 4. These photographs demonstrate the reduction in rotational-distortion when the electric field is rendered uniform:

a) with tube not fitted with gauze, and old type of coil;

b) tube with gauze, and coil as shown in fig. 12.

(In order that a direct observation might be taken, a fluorescent plate was mounted in each of the two tubes to take the place of the ordinary target. The fluorescent images produced on these plates are reproduced in the above photographs. The black circle and line in (a) are images of a figure applied to the photo-cathode for the purpose of measuring the magnification.)

It will be seen that the amount of rotational distortion inherent in the new design is very much less than before (fig. 4*b*). (It should be added that, even with the tube not equipped with gauze, less distortion results from the use of a coil similar to that designed for the tube with gauze. The results are, however, not quite as good as those obtained from the modified tube.)

Sensitivity

Lastly there is the question of the sensitivity of the new tube as compared with that of the tube without the gauze. The gauze unavoidably intercepts some of the photo-electrons, but a new process has made it possible to manufacture a very fine gauze with a ratio of aperture to material of 75 to 80%, which means that the loss in sensitivity is only some 25 to 20%.

Methods of varying the field of view

The production of T.V. programmes demands that at one moment a large part of a particular scene shall appear on the screen and at another only a small part of it; in other words that the field of view of the camera shall be variable. The means of achieving this may be either optical or electrical.

Optical means of varying the field of view

Optical variation is obtainable in the first instance by moving the camera closer to, or away from the scene. This facility, however, is effective only when the camera is already close to the scene, as is usually the case in the studio only.

The same effect can be produced by varying the focal length of the lens; the greater the focal length, the smaller the field.

This is explained in the following manner. Since the focal lengths in normal use are small compared with the object distance (camera to scene), the lens forms an image a very small distance beyond the focal plane, i.e. at a distance s' only slightly greater than the focal length f' (fig. 5). The field angle β is determined by:

$$\tan \beta = \frac{\frac{1}{2}d}{s'} \approx \frac{d}{2f'}$$

where d is the diagonal of the image. For a given value of d , therefore, a smaller field is obtained according as f' is increased.

In most T.V. cameras such variation in the focal length is obtained by means of a turret with 3 or 4 lenses of different focal lengths, a larger or smaller field being thus provided in steps, but a disadvantage of this system is that, when the lens is changed,

the picture momentarily disappears; the usual practice, accordingly, is to switch over to another camera in the meantime.

Generally it is then unavoidable that the second camera picks up the scene from another angle, and the sudden change in direction and range tends to disturb the continuity of the action as seen on the receiver.

A much better effect is obtained by varying the focal length continuously and, in fact, lens systems have been developed, known as Zoom lenses, which are capable of such variation. They are, however, very heavy and the cost is extremely high. One such lens system⁴) is 75 cm (29") in length and weighs 6.5 kg (14 lbs); another^{5,6}) 48 cm long (19"), weighs 18 kg (40 lbs).

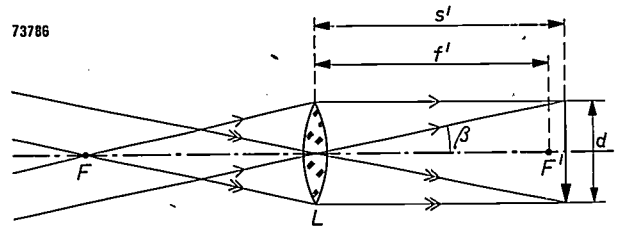


Fig. 5. The lens L , whose foci are F and F' , forms an image of a remote object at a distance s' which is slightly greater than the focal length f' . β is the angle of view.

Electrical means of varying the field

The potential pattern on the target of the tube is scanned by a beam of electrons to produce the video signal. If the horizontal and vertical amplitudes of the scanning beam are reduced, only part of the image is transmitted and the field of view is thus reduced in size.

The advantages of this system are that, with only simple apparatus, it provides continuous control of the field of view, and that it can be applied to all current types of camera tube. It is a disadvantage, however, that when the field of view is reduced, the definition drops, owing to the fact that the cross-section of the scanning beam is independent of the amplitude; when the area scanned is reduced, the definition suffers.

With camera tubes such as the image iconoscope, which are operated with a scanning beam of high velocity electrons, there is the additional disadvantage that the secondary-emission coefficient of the area scanned is also somewhat affected. This is manifested as a sort of target burn when the reduced

⁴) R. C. A. Broadcast Equipment Catalog 1950, p. 206.

⁵) J. Opt. Soc. Amer. 41, 863-864, 1951 (No. 11).

⁶) H. H. Hopkins, Television Zoom lens. Conv. British contribution to television, 1952, paper 1353.

scanning amplitudes are once more restored to their normal values, in that the rectangle originally scanned remains visible in the image.

In contrast with other camera tubes, image iconoscopes of the magnetic type — including the Philips model — offer facilities for varying the size of the field by other means, i.e. by varying the electron-optical magnification.

Electron-optical variation of the field of view

As already mentioned, a potential pattern 45 mm × 60 mm on the target of the Philips image iconoscope normally corresponds to an emission-image 12 mm × 16 mm on the photo-cathode; hence the electron-optical magnification M is 3.75.

If the magnification is greater, the edges of the photo-emission image are projected outside the scanned area of the target. In other words, the scanned (and transmitted) image then corresponds to part of the photo-emission image only; the field is thus smaller and the effect is the same as if the camera were moved closer to the subject. It should be added, however, that the resolving power of the electron lens, even with the gauze, is so high that with $M = 3.75$ it by no means limits the definition of the image; as mentioned in article I, the definition is limited by the electron gun. Hence the image on the target remains sufficiently clear, even when the magnification is increased.

As stated in article I (p. 128), any reduction in the magnification below 3.75 results in an increase in the optical errors, e.g. curvature of field, until a point is reached when they can no longer be compensated. For this reason we shall in the following consider only a magnification of 3.75 or more.

In our discussion of fig. 3a we remarked upon the possibility in principle of varying the magnification by axially displacing the focusing coil; when it is moved away from the target (which corresponds to a reduction in the effective length of the magnetic field) the magnification is increased. However, most cameras provide little opportunity for displacement in this direction owing to the fact that a short-focus objective is used and that the coil would quickly come to rest against the lens mounting. Movement of the coil is therefore only resorted to in order to adjust the magnification to a particular value once and for all. At the same time, the effective length of the magnetic field can also be varied by electrical means⁷⁾. Something in the shape of a

focusing coil in two sections might be visualised (fig. 6). If only that section which is at the photo-cathode end be energized, a short magnetic field will result and the magnification of the image will

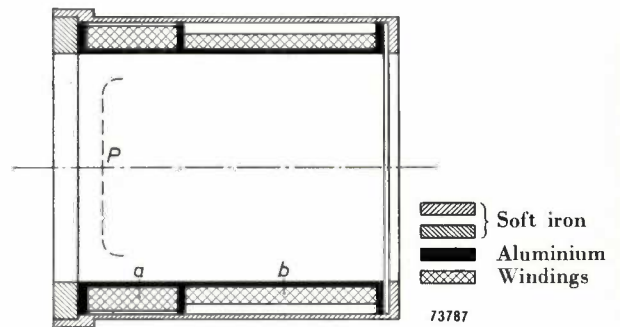


Fig. 6. Experimental focusing coil in two sections a and b , for continuous variation of the electron-optical magnification. P = relative position of the photo-cathode.

be higher than usual. With current applied to the other section, the current flowing in the first being simultaneously reduced to restore the focus, the image of the photo-cathode on the target will be smaller than before. The diagram in fig. 7 shows the magnification M plotted against the current I_2 flowing in the second coil, as well as the current I_1 in the first coil required to maintain a sharp image. These measurements were obtained from an image iconoscope with

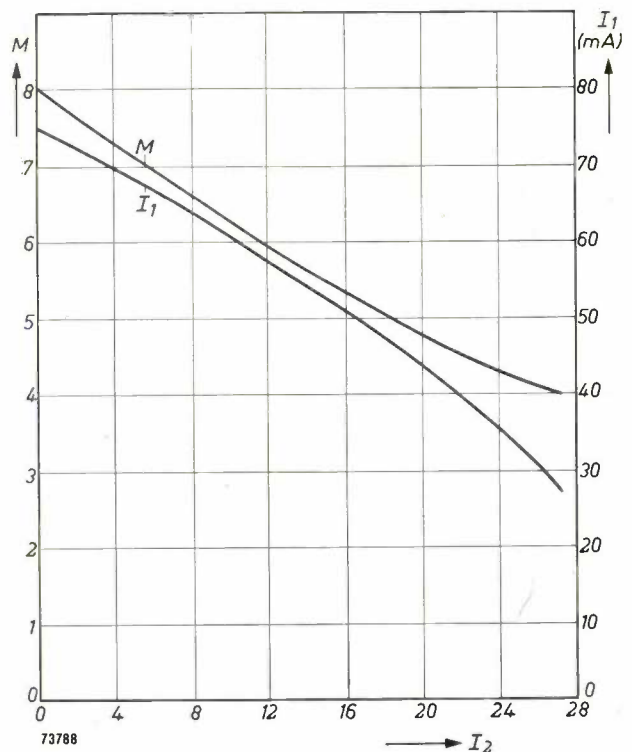


Fig. 7. The electron-optical magnification M as a function of the current I_2 flowing in section (b) of the coil shown in fig. 6. I_1 is the current that must flow in section (a) to ensure good definition.

⁷⁾ J. C. Francken., Tijdschrift Nederlands Radiogenootschap, 16, 243-257, 1951 (No. 5), in particular p. 252-256 (in Dutch); H. Bruining, Le Vide 7, 1248-1255, 1952 (No. 42) (in French).

gauze, in conjunction with the coil depicted in fig. 6. In this way, then, we have already achieved an appreciable variation in the magnification.

There is, however, a big disadvantage in this method. As outlined in article I, the electron-optical image is subject to a rotation of the potential pattern on the target with respect to the emission-image on the photo-cathode. It is found that when the magnification is varied in the manner just described, the angle of this rotation varies as well, which is most undesirable.

It would be possible to make some compensation for this rotation by turning the tube an equal amount in the opposite direction inside the camera, but, obviously, this would be a makeshift and complicated arrangement. By ascertaining the factors governing the angle of rotation it has been found possible to evolve a design in which the angle remains constant while the magnification is varied.

Relationship between magnification and angle of rotation

The following experiments have thrown more light on the dependence of the rotation of the image on the form of the magnetic field.

Use was made of the kind of focusing coil depicted in fig. 8. This consists of two sections enclosed in a

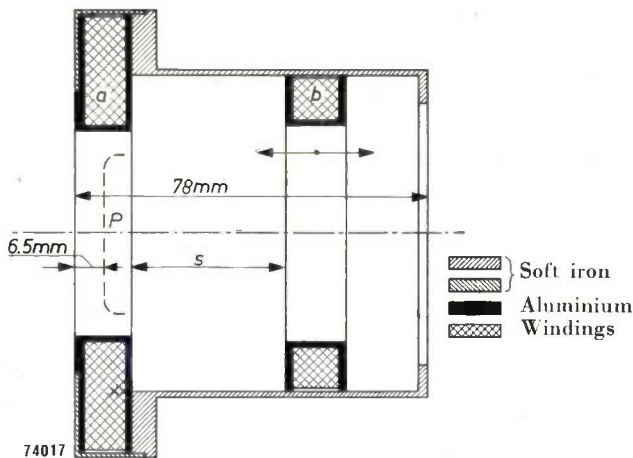


Fig. 8. Experimental focusing coil with fixed section (a) and adjustable section (b). *P* = relative position of the photo-cathode.

soft iron housing; the section facing the target is movable. The experiments were carried out on a tube which had been provided with a gauze. Very similar results were obtained, however, from a tube without a gauze.

The angle of rotation Ψ was determined as a function of the magnification M , for several

different positions of the movable section of the coil. The results of some of the measurements are shown graphically in fig. 9. The parameter in this figure is the distance s between the two coil sections. With

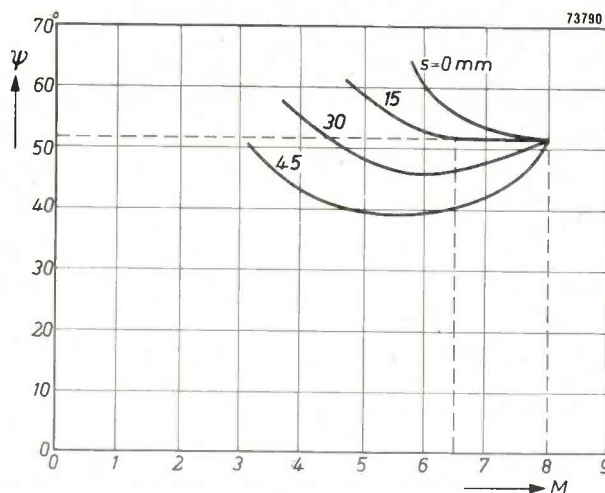


Fig. 9. The angle Ψ through which the image on the target is rotated with respect to the image on the photo-cathode, as a function of the electron-optical magnification M . The parameter varied here is the distance s between the two sections of the coil (fig. 8).

maximum magnification ($M = 8$) no current flows in coil section *b*, so that Ψ is then independent of s , that is, the curves meet at a common point.

The following conclusions may be drawn from the chart:

- 1) A particular magnification can be obtained with magnetic fields of different forms, but the angles of rotation then differ: the straight line $M = \text{constant}$ is intersected by the curves $\Psi = f(M)$ at various points, which means that the angles differ.
- 2) The slope of the curves in the region of $M = M_{\text{max}}$ is negative when the second coil section is close to the first, but becomes more and more positive according as it is moved in the direction of the target.
- 3) The angle of rotation has a minimum in those cases where the curves have positive slopes at $M = M_{\text{max}}$.

It follows from conclusion 2) that the angle of rotation in the vicinity of $M = M_{\text{max}}$ can be kept constant by finding the appropriate position of the second coil section (see curve for $s = 15$ mm). The magnification range extends only from $M = 8$ to $M = 6.5$. In order to ensure a constant angle between more widely separated limits of the magnification, the second coil section in fig. 8 has to be divided in two (fig. 10); the two sections thus obtained must then be so adjusted that one of them

produces a positive slope of $\Psi = f(M)$ near $M = M_{max}$, and the other a negative slope, so that the two effects compensate each other. It has actually been found possible in this way to maintain a con-

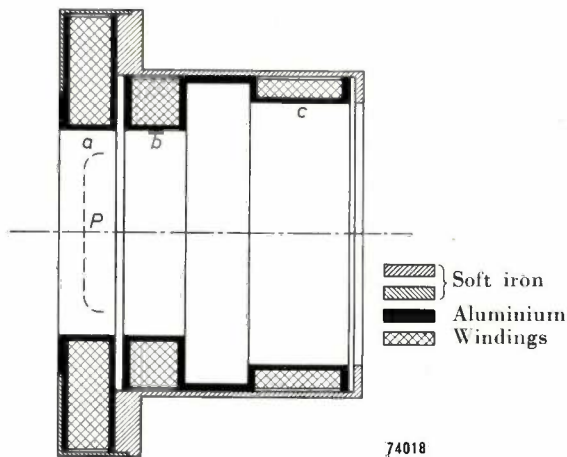


Fig. 10. Focusing coil with three fixed sections (a), (b) and (c) which can be separately energized. P = relative position of the photo-cathode.

stant angle of rotation between wide limits of the magnification, and fig. 10 shows the arrangement for a coil which will allow of variation in the magnification in the ratio of 1 : 2, with constant angle of rotation.

Fig. 11 illustrates the current in the three coil sections plotted against the magnification. For continuous variation of the magnification these currents must be so adjusted, simultaneously, that the curves as shown in this figure are reproduced. The method employed to achieve this is described in the last section of this article.

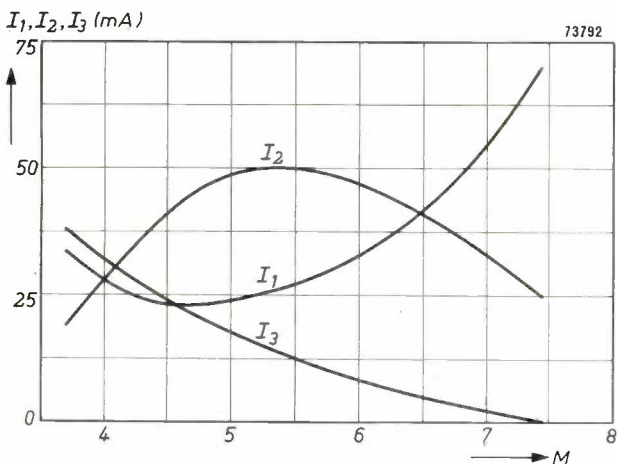


Fig. 11. The currents I_1 , I_2 and I_3 which must flow in the three sections of the focusing coil (Fig. 10) to ensure high definition and almost constant angle of rotation Ψ plotted against the magnification M . M can be varied continuously in the ratio of 1 : 2.

The magnetic field of the focusing coil protrudes to a certain extent in the space within which the scanning beam moves. The beam cuts the magnetic lines of force almost at right angles and the extent to which the beam is consequently deflected is more marked at the top of the target than at the bottom, the scanned figure being accordingly a parallelogram instead of a rectangle (diamond distortion). This may be corrected by placing a small coil between the focusing coil and the target to compensate locally the field of the focusing coil. The current flowing in this small coil would have to be proportional to the current in the third section of the focusing coil, which contributes by far the most towards the field in the scanning space. The compensating coil, shunted by a resistor which would be adjustable and subsequently locked permanently, would accordingly be placed in series with the third coil section. In figs. 13 and 14 this is not shown.

We may add here that a coil in three sections (fig. 10) proved to be a very suitable basis for the design of a coil to give a fixed magnification (3.75 \times) in the image iconoscope fitted with gauze. By energizing the three sections in different ways we have been able to find the optimum image with the desired magnification and, from the values of the

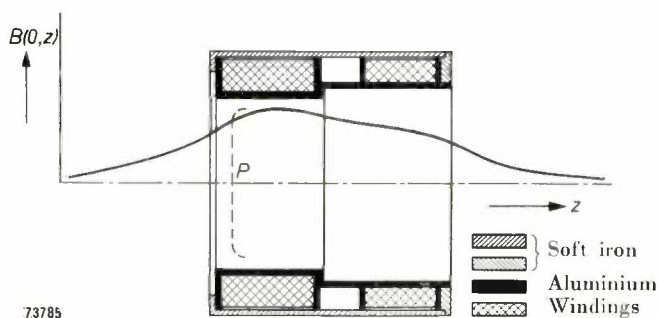


Fig. 12. Focusing coil with two sections in series to give constant magnification ($M = 3.75$) in the image iconoscope with gauze. P = relative position of the photo-cathode.

current in the three coil sections it was finally found possible to work out details for a single coil having different numbers of turns per centimetre in the axial direction. A sketch of such a coil is shown in fig. 12; the form of the resultant field is such that the image exhibits little or no distortion, and only traces of curvature.

Practical application of the variable magnification

In order to incorporate the variable magnification described above in a T.V. camera, it is necessary to adjust the currents flowing in the three sections of the magnetic lens in accordance with the curves depicted in fig. 11. For this purpose each coil can be connected to a potentiometer (fig. 13). In practice it is essential that the three potentiometers be operated by a single control. For the first section an

ordinary linear potentiometer can be used; the two others should conform to special laws such that the values of these currents I_2 and I_3 will suit each value of I_1 in accordance with fig. 11.

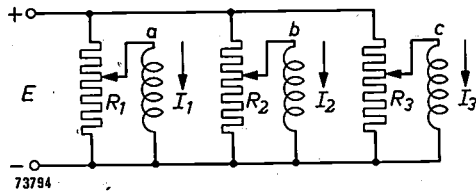


Fig. 13. Circuit for simultaneous control of the currents I_1 , I_2 and I_3 in the three sections a , b and c of the focusing coil. E = constant control voltage. R_1 , R_2 , R_3 = variable potentiometers.

A drawback of this arrangement is that the resistance values of the coil sections, as well as those of the potentiometers, tend to vary with changes in temperature. For this reason it is better to connect a high-impedance electron tube in series with each

devised a special circuit for the potentiometers R_1 and R_2 (see fig. 14). R_1 has a fixed tapping connected to the negative line, the two extremities being taken to the positive; R_2 also has a fixed tapping, but this is connected to the positive line, with both ends of the track to H.T. negative. When the control knob operating R_1 , R_2 and R_3 is so adjusted that the slider of R_1 coincides with the tapping point, tube P_1 receives maximum negative bias, i.e. I_1 is at its minimum. Rotation of the knob one way or the other causes I_1 to rise. When the adjustment is such that the slider of R_2 lies on the tapping point, tube P_2 receives minimum negative bias and I_2 is at a maximum.

Finally, a few remarks about the sensitivity of the camera tube when operated with variable magnification. It will be obvious that when a change is made from an effective area of the photocathode of 20 mm diagonal to another having a diagonal of only 10 mm, the average effective photo-current for that particular area will be smaller. The actual decrease will depend on the amount of information contained in the scene being televised, but, when the brightness of the scene is

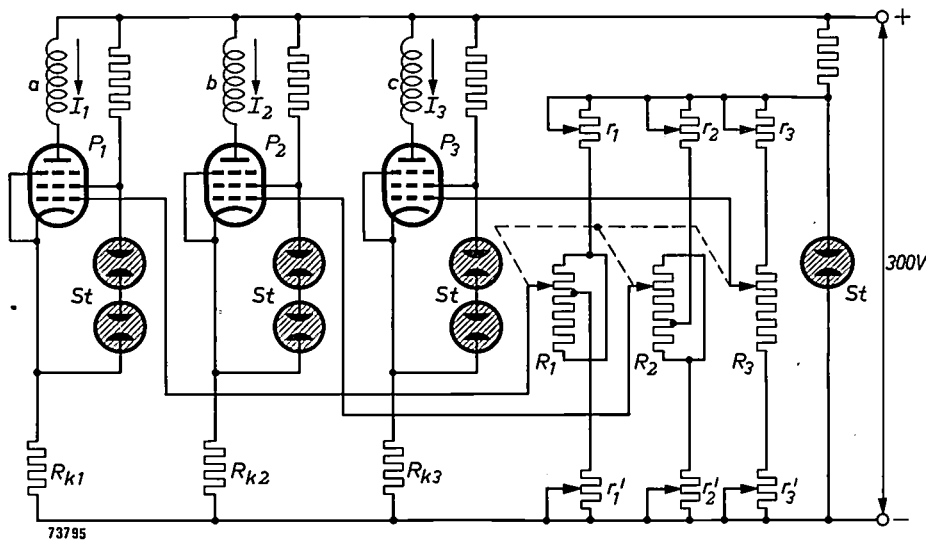


Fig. 14. Improved circuit for the simultaneous control of currents I_1 , I_2 and I_3 in the three sections a , b and c of the focusing coil. P_1 , P_2 and P_3 each denote two pentodes type PL 83 connected in parallel. St = stabiliser tubes type 85-A1. R_1 , R_2 , R_3 = potentiometers operated by a single control knob. r_1 , r_2 , r_3 and r'_1 , r'_2 , r'_3 = correcting resistors. R_{k1} , R_{k2} and R_{k3} = biasing resistors producing strong negative feed-back so that the anode currents shall be independent of the valve characteristics.

coil. The grid bias of each of the three tubes can then be controlled so as to provide an anode current of the required value. A suitable circuit ⁸⁾ is shown in fig. 14.

For the purposes of fig. 13 it was tacitly understood that I_1 , I_2 and I_3 were monotonically rising or falling functions of M but, in fact, as will be seen from fig. 11, only I_3 is a descending curve, whereas I_1 exhibits a minimum and I_2 a maximum. In order to be able to control these three currents in the desired manner with a single control knob we have, therefore,

⁸⁾ Designed by A. J. Sietsma of this laboratory.

roughly uniform, the average photo-current will drop to 1/4 of its original value. From fig. 18 in article I it appears that the signal delivered by the tube does decrease. The result is a dimmer image.

There are two ways of maintaining a constant level of video signal. The first of these consists of ensuring by optical means that the illumination on the photocathode varies in inverse proportion to the effective area of the cathode. One way of doing this is to use larger stops as the effective area of the photo-

cathode is reduced. The other method is based on electrical compensation, viz. an increase in the amplification factor of the video amplifier when the magnification of the camera tube is increased.

Clearly, neither of these two methods can be carried very far. In the first instance very fast lenses are used — even with a photo-cathode of 20 mm diameter — and only a limited use can be made of the stop; in the second instance too much amplification results in too much background noise. This means that with the camera tube type 5854/02 we are restricted, in the matter of variation of the magnification, to a factor of 2.

Another limitation on the use of very small areas of the photo-cathode is imposed by the resolving power of the lens, as will be clear when we compare the case of a photo-cathode of which full use is made, with one that is used only in part; in both cases the definition (the number of lines in the height of the image) must satisfy the same minimum requirement. The resolving power of the lens (number of lines per centimetre that can be resolved) has therefore to conform to the highest standard for the cathode when only part of this is used. In addition, the stop used in the latter instance has to be larger (see above).

On the basis of very recent experiments it has been found possible to design a soft iron housing for

the coil, such that the maximum diameter of the effective part of the photo-cathode can be increased from 20 mm to 25 mm, enabling the magnification to be adjusted up to a ratio of 1 : 2.5 without the disadvantages mentioned above becoming apparent.

Summary. In image iconoscopes positive ions are produced which migrate towards the photo-cathode. In the older type of tube the non-uniform electric field concentrates these ions upon a small area of the photo-cathode with the result that this area rapidly loses sensitivity (ion burn or ion spot). In the new Philips image iconoscope (type 5854/02) a fine metal gauze is mounted at the end of the anode cylinder, facing the photo-cathode. Between this gauze and the cathode the electric field is uniform, the positive ions are accordingly distributed uniformly over the whole area of the photo-cathode and the occurrence of an ion spot is eliminated.

Provision of the gauze brings with it another advantage, namely a reduction in the (already small) amount of rotational distortion in the electron-optical image of the photo-cathode on the target. As against these advantages there is but one drawback, a slight reduction (20 to 25%) in the sensitivity, as a result of interception of electrons by the gauze.

A second improvement in the image iconoscope concerns the control on the field of view, which is now variable in size by wholly electrical means, necessitating no complicated optical systems. For this purpose the magnetic focusing coil is made in three sections, which are energized separately. The three currents can be controlled simultaneously, so that the electron-optical magnification, and with it the field of view, can be varied continuously in the ratio of 1 : 2. At the same time the image remains sharp and the angle through which the image is rotated with respect to that on the photo-cathode remains practically constant. A circuit is described which enables the energizing currents for the three sections of the focusing coil to be adjusted in the desired manner by means of a single control knob.

A RUBBER MEMBRANE MODEL FOR TRACING ELECTRON PATHS IN SPACE CHARGE FIELDS

by G. A. ALMA, G. DIEMER and H. GROENDIJK. 537.533.3.072: 621.385.1

The rubber sheet has for some time been a well-known medium for the investigation of two-dimensional fields and the motion of electrons within them. It has been found possible in the development of electron tubes by this means to obtain a better insight into many problems, in particular in cases where purely theoretical treatment presents great difficulty. The method is also useful when applied to ultra-short-wave valves, where the electrode spacing is extremely small, provided that allowance be made on the rubber sheet for the space charge produced by the electrons themselves. A practical method of doing this, and the principles on which the method is based, are described in the following.

The motion of the electrons in electron tubes can be imitated by rolling steel balls over a surface which is a "model" of the electrostatic field in the valve; a stretched rubber sheet forms a conveniently deformable surface for the purpose. The vertical distance h of such a surface from a fixed horizontal surface is at every point (X, Y) proportional to the potential φ at the corresponding points (x, y) in the field. Furthermore, the horizontal projections of the tracks of the steel balls in the model are of the same form as those of the electrons in the electrostatic field; the velocities of the balls and electrons are also proportional. The proof of this may be found in an article by Kleijnen which appeared in this Review

some time ago¹⁾ and which will be referred to in the following as I.

The paths and velocities of the brightly polished steel balls are determined by photographing them with intermittent light of known periodicity; in this way these paths are rendered visible as dotted lines (fig. 1), the velocity being ascertained from the spacing of the dots.

Imitating the space charge by exerting pressure on the rubber sheet

In his work Kleijnen assumed that the electrostatic field did not include a space charge (see I, pp. 338-339), but we shall extend our considerations to make allowance for a space charge. We proceed from the assumption that the rubber sheet is first stretched with a constant tension σ in a horizontal plane XY . We then consider the situation when the sheet is displaced vertically; the vertical deflection h of any point will be a function of its coordinates X and Y and, in the general case, also of the time t . It is assumed, further, that a vertical pressure p operates on the sheet, which in general will also be dependent on X, Y and t . The tension of the sheet (see I) results in a vertical force on an element of dimensions dX and dY equal to:

$$\sigma \left(\frac{\partial^2 h}{\partial X^2} + \frac{\partial^2 h}{\partial Y^2} \right) dX dY.$$

The pressure results in a vertical force $p dX dY$. Let m_1 denote the mass per unit area; the equation of motion will then be:

$$\sigma \left(\frac{\partial^2 h}{\partial X^2} + \frac{\partial^2 h}{\partial Y^2} \right) + p = m_1 \frac{\partial^2 h}{\partial t^2} \quad (1)$$

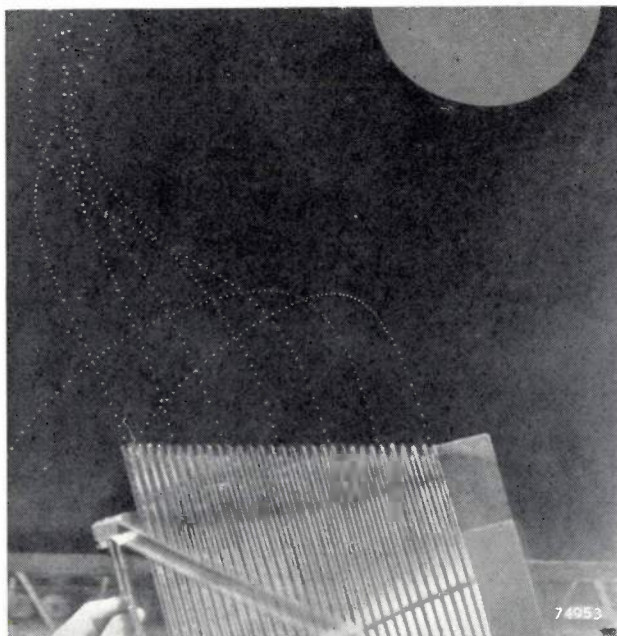


Fig. 1. The movement of electrons in an electrostatic field can be imitated by rolling steel balls over a stretched and appropriately supported rubber sheet. When photographed with intermittent light the tracks of the polished steel balls are seen as dotted lines. The velocity can be ascertained from the spacing of the dots.

¹⁾ P. H. J. A. Kleijnen, The motion of an electron in two-dimensional electrostatic fields, Philips tech. Rev. 2, 338-345, 1937, in particular page 340. The method can be employed only when the field is independent of one of the three mutually perpendicular coordinates.

We are interested only in the case where the rubber sheet is in equilibrium²⁾; then $\partial^2 h / \partial t^2 = 0$ and p is independent of time, so that equation (1) is reduced to:

$$\frac{\partial^2 h}{\partial X^2} + \frac{\partial^2 h}{\partial Y^2} = -\frac{p}{\sigma} \dots (2)$$

This is Poisson's equation, which is also valid for the potential φ in electrostatic fields with space charge. If the density of the space charge be denoted by ρ , we may write for this potential in a two-dimensional field (co-ordinates x and y):

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = -\frac{\rho}{\epsilon_0} \dots (3)$$

where $\epsilon_0 = 8.855 \times 10^{-12}$ A.sec/V.m, the dielectric constant of free space.

The rubber membrane is prepared in such a way that the following relationships exist between the co-ordinates of corresponding points in the field and model:

$$\left. \begin{aligned} x &= S_l X \\ y &= S_l Y \end{aligned} \right\} \dots (4)$$

where S_l is a scale factor for the length dimension. It is also required that the vertical displacement of the rubber sheet and the potential of the field at corresponding points are related as:

$$\varphi = S_\varphi h \dots (5)$$

where S_φ is the scale factor for the potential.

By applying (4) and (5) to (3) and then comparing this differential equation with (2), we find that p should conform to:

$$p = \frac{S_l^2}{S_\varphi} \frac{\sigma}{\epsilon_0} \rho \dots (6)$$

The solution of Poisson's equation (2) or (3) is uniquely determined if the right-hand term is given as a function of the co-ordinates and provided that the boundary conditions are known. In this case the boundary conditions are that the potentials at the surfaces of the electrodes should have specified values. In order to obtain the desired model

of the field, therefore, we must apply a pressure at every point on the rubber sheet that will be in accordance with (6); the mechanism employed to ensure this is described at a later stage. Moreover, we must produce deflections h along the lines corresponding to the edges of the electrodes, that will be related to the potentials of these electrodes in accordance with (5).

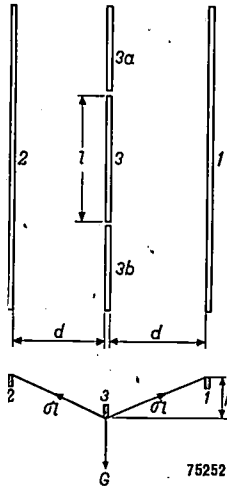


Fig. 2. Method of determining the tension in the rubber sheet. The parallel rods 1 and 2 are mounted horizontally and the rubber sheet is laid over them. Rod 3 (length l) is depressed with a known force G ; the sag h resulting from this force is then measured. The additional rods 3a and 3b, which are pressed down to the same depth as rod 3, serve to counteract the effect of the finite length of rod 3. From the lower figure, which is also the diagram of forces, it is found that when $h \ll d$, $G = 2 \sigma l h / d$, from which it follows that $\sigma = G d / 2 l h$.

To be in a position to employ equation (6) we must apparently know the tension σ of the sheet; this is determined experimentally by measurement in the manner shown in fig. 2 and described in the caption to the figure.

Method of successive approximation

As will be apparent from the foregoing, the space charge distribution must be known in order that the model can be made.

Apart from the number of electrons, however, the space charge depends on their paths and velocities, which are just those unknowns that have to be determined from the model. A way out of this difficulty is offered by the method of successive approximation. A model is first made without space charge (pressure), or still better, with a pressure corresponding to a plausible space charge distribution; from the motion of the steel balls then observed, the corresponding density of the space charge can be computed³⁾ and the distribution

²⁾ In the absence of equilibrium and assuming $p = 0$ (which means that there is no pressure on the rubber sheet), equation (1) takes the form:

$$\frac{\partial^2 h}{\partial X^2} + \frac{\partial^2 h}{\partial Y^2} = \frac{m_1}{\sigma} \frac{\partial^2 h}{\partial t^2}$$

This is known as the wave equation, and describes transverse vibrations of the rubber sheet. These vibrations are analogous to the electromagnetic waves occurring in cavity resonators. This analogy has been used in this laboratory in a study of the performance of cavity resonators; see K. S. Knol and G. Diemer. Philips tech. Rev. 11. 156-163, 1949.

³⁾ The contribution of an electron to the average space charge in a given volume element is proportional to the time that the electron dwells in this element. This time is in turn proportional to the number of dots produced on the photograph by the steel ball when traversing the corresponding surface element of the rubber sheet. Intuitively it may be said that the density of the space charge can be derived from that of the dots in the photograph, but further discussion of this point would take us beyond the scope of the present article.

of pressure modified accordingly. The steel balls are then again rolled over the sheet, the space charge distribution being again computed, and so on.

In general this procedure may be expected to yield an increasingly closer approximation to the correct solution, and a number of simple examples in which the method of successive approximation can be verified mathematically have shown that this can, in fact, be achieved quite quickly. Nevertheless, in certain instances the convergence may prove to be too slow, or there may actually be divergence (see Appendix).

Some experiments, which are described in the following, confirm the assumption that rapid convergence usually does take place⁴).

Experimental equipment

Charged particles are repulsed by a space charge of their own polarity. When they enter a zone in which such a space charge exists, their potential, therefore, is greater than in the absence of the space charge; hence, in the model, an upward pressure on the rubber sheet will correspond to a space charge. This upward thrust is provided by a number of pads to which a vertical force is applied, the amount of force per pad being variable. The forces are supplied by weights and are applied by means of wires and pulleys. A number of units (each having six such pressure pads, *fig. 3*) can be used at a time, this being a very convenient method, capable of representing a wide range of electrode configurations.

To minimize possible irregularities in the surface of the rubber sheet at the points where the boundaries of the pressure pads occur, the latter are made of sponge rubber and are mounted in such a way that they will tilt easily.

⁴) Mention has already been made in I of the method of successive approximation as a possible means of taking into account the space charge caused by the electrons themselves. It is however only when this process is used in conjunction with the method of imitating the space charge by applying pressure to the rubber sheet that the process can be used successfully for practical purposes. In principle, it is possible by applying a numerical solution of Poisson's equation, to compute the correction that has to be made in the depth of the rubber sheet to allow for the space charge, but this entails an enormous amount of work. Moreover, after the first stage, the rubber sheet would have to be replaced by a series of rigid computed surfaces which, of course, would be impracticable. The idea of representing the space charge by a variable pressure has in the meantime also been suggested by Bloch in a discussion subsequent to a lecture on the errors in the rubber sheet (*Proc. Inst. El. Engrs.* 97, II, 443, 1950, and also independently by the Russian scientists Bobykin, Kelman and Kaminsky, who have made practical use of this method (*J. tech. Phys. USSR*, 22, 736-743, 1952). They employed apparatus very similar to ours (and also applied the pressure by means of plates loaded with weights).

It is essential that the forces as adjusted shall operate regardless of the height ultimately assumed by the rubber sheet, and the unit described fulfils this condition. Pressure applied by means of electromagnets, a method which was tried and which has the advantage of simple remote control, does not sufficiently meet this requirement.

Owing to friction in the apparatus, the forces applied to the sheet are not strictly in accordance with the loading weights. It is possible to introduce a correction for this, but friction should in any case be kept to a minimum. In the unit shown in *Fig. 3*, the friction is less than 5% of the weight applied. For some of the experiments use was made of another, less efficient mechanism, in which friction accounted for 15% of the applied weight. Reference to the method of compensating the friction losses will be made later.

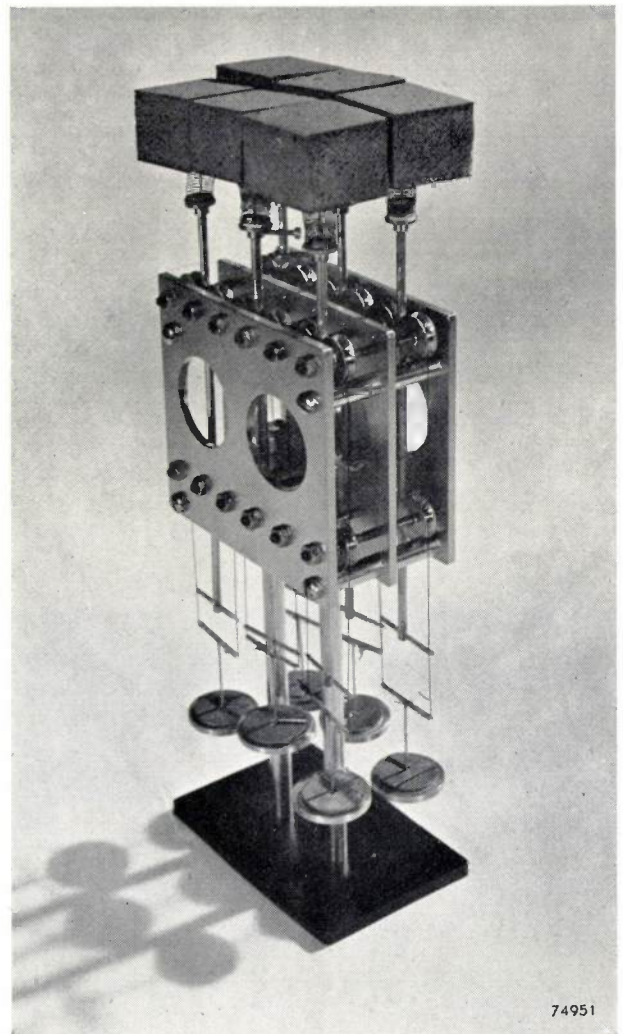


Fig. 3. Unit for applying upward pressure to the rubber sheet. This pressure represents the space charge produced by the electrons. A number of such units placed side by side can be used to imitate any distribution of space charge. The dimensions of the pressure pads are 3 cm x 3 cm.

Experiments to test the method

In order to test the method three sets of conditions were imitated on the rubber sheet, relating to configurations in which the electron paths were already known; tests carried out with the rubber sheet yielded these results with a very reasonable degree of accuracy, thus considerably adding to our confidence in the method.

Rubber sheet shaped cylindrically

The first experiment described was intended to examine the effect of the separate pressure pads on the paths of the steel balls: it was feared that irregularities might appear in the surface of the sheet at the transitions between adjacent pads. Such discrepancies would have represented a departure from the theoretical paths and, especially with low velocities, this might be quite pronounced.

The apparatus used for this test is depicted diagrammatically in fig. 4a. The rubber sheet was curved in one direction only, namely the X-direction the trough being obtained by placing the sheet on

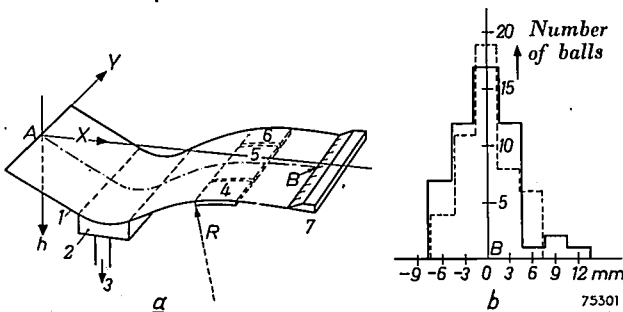


Fig. 4. a) Diagram illustrating the method employed for checking the effect of the arrangement of pressure pads in the space charge unit (fig. 3) upon the tracks of the steel balls. 1 rubber sheet, 2 box with perforated upper surface, connected to vacuum pump 3; 4, 5 and 6 cylindrical metal plates imparting to the rubber sheet a curvature of radius R . Plate 5 can be replaced by the space charge unit shown in fig. 3, 7 rule.

b) Deviation from the point B , of the points of arrival of steel balls started at A . The ordinate indicates the number of balls arriving within the distance shown by the abscissa. The full line refers to the rubber sheet as supported by the metal plate 5, and the broken line to the case where the plate is replaced by the space charge unit.

a metal box, the hollowed top of which was perforated, so that, with the box connected to a vacuum service line, the sheet would be firmly held in position. Further along, the sheet was restored to the horizontal by a smooth, cylindrically curved plate of radius R . This plate was made in three pieces, as shown by the dotted lines in fig. 4a, the dimensions being such that, with the centre part removed, there would be just enough room for a pressure unit with six plates.

Under these conditions the theoretical values are easily calculated for the forces with which the pressure plates should be loaded in order to reproduce as faithfully as possible the contour of the smooth metal plate. It was found that the necessary pressure on the whole of the supported surface must be constant at $p = \sigma/R$. A check was made to see whether these computed loads would yield the desired contour without serious irregularities, and for this purpose steel balls were started from the point A (Fig. 4a), without initial velocity; in the ideal case these should all have arrived at the point B , lying opposite to point A in a direction perpendicular to the "electrodes". In actual fact, with both the smooth plate and the "space charge apparatus", a certain amount of spread occurred at the point of arrival, and this was measured by means of a rule placed at B . The number of balls arriving within certain limits is shown in the chart in Fig. 4b. Although the velocity at and after passing the supporting plate of radius R was low, so that the effect of irregularities in the surface of the rubber sheet might well have been quite pronounced, the diagram shows that the deviations were very similar in both cases. Apparently, then, differences caused by imperfections in the space charge model are only slight compared with those due to other causes.

Pierce electron gun

As an instance intended to demonstrate the speed with which the method of successive approximations yields results, a form of the well-known Pierce electron gun⁵⁾ was reconstructed. In this type of electron gun the form of the electrodes is such that the electrons travel from the cathode to the anode in a parallel beam, despite their natural repulsion (fig. 5). When no pressure was applied to the rubber

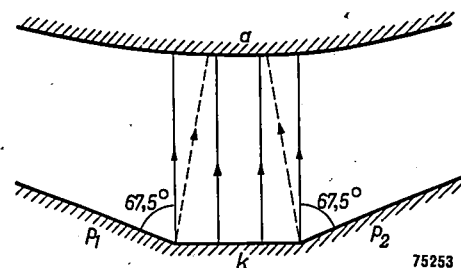


Fig. 5. Electrode configuration of the Pierce electron gun. k cathode; p_1 and p_2 auxiliary electrodes at cathode potential; a anode. On the rubber sheet the "electron paths" were within the broken lines, when no space charge (i.e. distributed pressure) was applied. With the "space charge" as derived from these paths the tracks at once became parallel.

⁵⁾ J. R. Pierce, Rectilinear electron flow in beams, J. appl. Phys. 11, 548-554, 1940; see also Philips tech. Rev. 13, 216, 1952.

sheet (i.e. without space charge), the beam was found to converge to such an extent that, on arrival at the anode, it had roughly only half its original cross-section. When the density of the space charge was computed as a first approximation from the tracks obtained, pressure being then applied to simulate the corresponding distribution, the beam proved to be parallel within the limits of accuracy obtainable with the rubber sheet ⁶⁾.

This result was obtained in spite of the fact that the first approximation of the space charge distribution differed considerably from the actual distribution, which was known from Pierce's calculations ⁵⁾. A second approximation of the space charge could have been computed on the basis of the parallel beam obtained from the first approximation, and this would have been very close indeed to the actual conditions.

This example shows that the distribution of pressure as applied to the rubber sheet may deviate considerably from the true value, but that for practical purposes the balls nevertheless simulate very closely the actual electron paths.

Diode with thermal velocity distribution of the electrons

The case of the plane diode can be treated theoretically even if account is to be taken of the different initial velocities with which electrons are emitted from the cathode (thermal velocity distribution). In a plane diode, velocity components parallel to the electrodes have no effect on the transit time of the electrons, or on the distribution of the space charge and potential. In principle, therefore, we need not make any allowance for the directional distribution of the initial velocities. Nevertheless this has been done, with a view to the intended investigation of a triode in which the problem is now two-dimensional owing to the presence of the grid wires; here the influence of the directional distribution of the electrons becomes important.

The thermal distribution of the velocity in the direction perpendicular to the cathode is depicted in *fig. 6*. As it is obviously impossible to reproduce such a continuous distribution of velocities on the rubber sheet, this was approximated by a distribution in four velocity groups of equal intensity, that is with the same number of electrons in each group. In *fig. 6* this division is represented by four zones of equal area, I to IV, in the space below the curve. Each group was imitated on the rubber sheet by five steel balls, started from the cathode

⁶⁾ Bobykin carried out the same check on the method, with the same results.

normally and at angles of 30° and 60° from either side of the normal, all with the normal velocity component appropriate to the group ⁷⁾. The dependence of the emission on its direction,

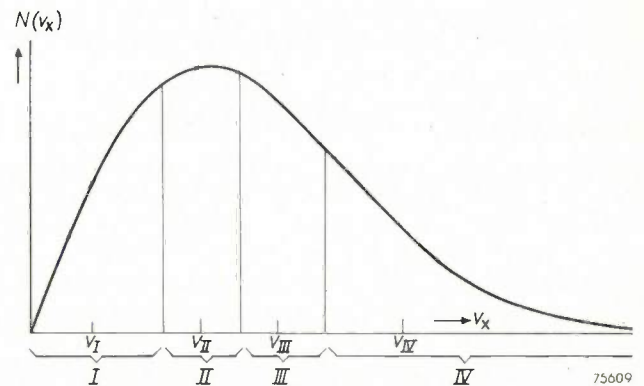


Fig. 6. Distribution of the forward velocity V_x of the electrons emitted by a heated cathode. The distribution is approximated to by division into four velocity groups, $V_I \dots V_{IV}$.

which is known ⁸⁾, was taken into account in determining the space charge distribution from the dotted tracks in the photographs, by multiplying the dot density by the appropriate factors. To smooth out irregularities in the rubber sheet and so secure a better average, the experiment was repeated for different starting points at the

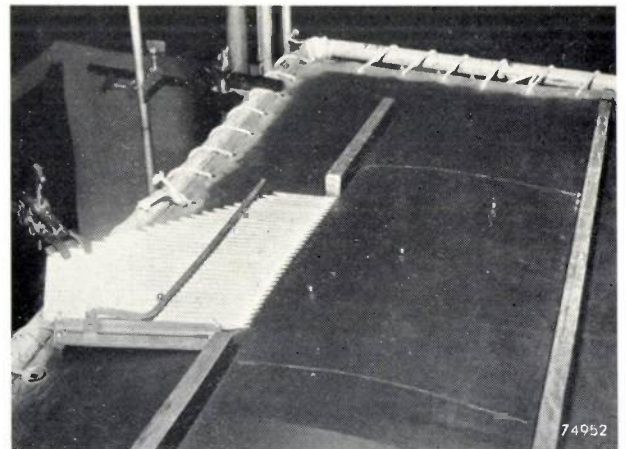


Fig. 7. View of the rubber sheet used for imitating the potential field of a diode. Two chalk lines have been made on the rubber sheet to demonstrate the curvature. At the left will be seen the starting chute used for imparting to the steel balls their initial velocity. In all the tests carried out, the balls were launched with sufficiently long intervals to prevent collision and too much indentation of the rubber sheet due to their combined weight.

⁷⁾ The arithmetic mean velocity of the electrons in a velocity group must not be used in this case. The contribution of an electron to a space charge in a volume element is proportional to the period that the electron dwells in the volume element, i.e. to the reciprocal of the velocity. For each group, therefore, the "appropriate" mean velocity is such that its reciprocal is the mean of the reciprocals for the electrons in that group.

⁸⁾ See amongst others, W. G. Dow. *Fundamentals of Engineering Electronics*, Chapman and Hall, London, 1937, Ch. XI.

cathode. Fig. 7 is a photograph of the rubber sheet used for reproducing the potential curve of the diode.

In this model a check was made to ensure that the sheet would assume the correct shape when pressure was applied in accordance with the exact space charge as computed from (6). The form of the sheet is depicted in fig. 8. The upper curve was obtained

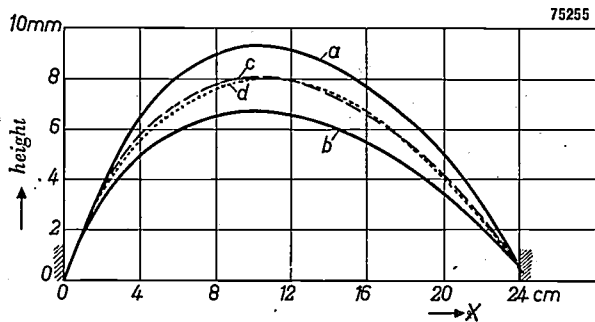


Fig. 8. Influence of the friction in the space-charge unit which was the forerunner of the one depicted in fig. 3. The potential curve of a diode was imitated on the rubber sheet (see fig. 7). In the diagram the height of the sheet in the cross-section perpendicular to the electrodes is plotted. Curves *a* and *b* represent the extreme positions of the rubber sheet that it would assume as a result of the friction. Curve *c*, the mean curve of *a* and *b*, almost coincides with the theoretical curve *d*.

after the sheet had been allowed to settle slowly on the pressure pads, the lower after pushing the sheet down further than necessary and allowing it to rise. The mean of the two contours so obtained corresponds closely to the theoretical potential distribution. From the deviations shown by the extreme curvature compared with the average, it is seen that the friction in the space-charge apparatus represents a certain percentage (in this case 15%) of the load. In our experiments the rubber sheet was each time lowered very carefully on the pressure pads, so that all frictional forces were directed upwards, and a correction was made for these forces by making the load on the plates 15% less than the value provided by equation (6).

With the rubber sheet thus correctly shaped the steel balls were rolled with the above mentioned velocities and directions, and the space charge distribution was again computed from the paths so obtained. This should have agreed with the theoretical distribution taken as the starting point and, in fact, this was found to be very nearly the case (fig. 9). The difference between the true curve (*a*) and that obtained (*b*) may be attributed to the method employed in compensating the friction of the balls on the sheet⁹⁾, namely by increasing

⁹⁾ This friction should not be confused with that which is inherent in the space charge apparatus, to which reference has been made.

their initial velocity. This is increased by an amount such that the correct velocity would be attained at the potential minimum; the electrons passing to the anode accordingly move too rapidly before reaching the potential minimum and, in that region, contribute too little towards the density of the space charge. Beyond the minimum they move too slowly and contribute too much.

Curves *c* and *d* in fig. 9 give some idea of the rate of convergence. A start was made on the basis of a space charge density corresponding to curve *c*, which at every point was twice the theoretical value and therefore incorrect. By rolling the balls and deducing the space charge distribution from the result, curve *d* was obtained, the average of which differs by less than 10% from the final result (curve *b*). When a start was made with a space charge equal to only one half of the theoretical value, the results also converged reasonably quickly (curves *e* and *f*). In both cases the error in the region of the potential minimum was quite small, even after the first stage of the process. Fig. 10

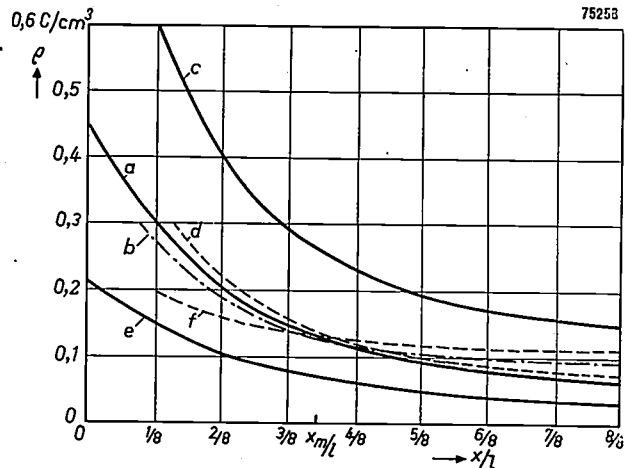


Fig. 9. Space charge density in a diode, as a function of the position co-ordinate x/l (l is the distance from cathode to anode). x_m/l represents the position of the potential minimum.

Curve *a*. Theoretical distribution of the space charge.

Curve *b*. Derived from tests on the rubber sheet with pressure applied corresponding to the theoretical space charge distribution.

Curves *c* and *e* represent distributions with the space charge at all points twice, and one-half of the theoretical values respectively. Curves *d* and *f* were derived experimentally on the rubber sheet with a pressure corresponding to the space charge distribution of *c* and *e* respectively. Curves *c*, *d* and *e*, *f* give an idea of the rate of convergence of the method of successive approximation.

shows the track of the balls as drawn from a photograph taken with intermittent lighting. These tracks refer to electrons starting at right angles to the cathode at three different velocities.

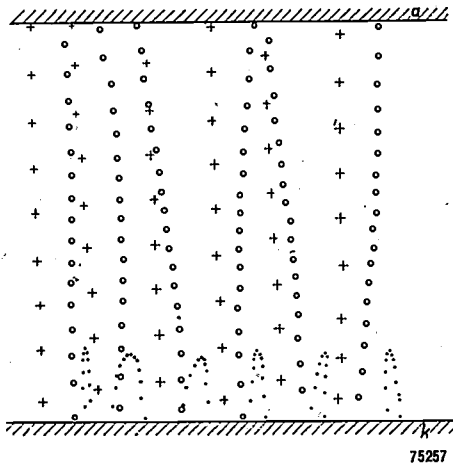


Fig. 10. "Electron paths" on the rubber sheet with which the potential field in a diode was imitated. The dotted tracks were drawn from photographs taken with intermittent lighting. Three different velocities were employed perpendicular to the cathode, these being represented by different symbols, viz. $0.84 v_m$ (\bullet), $1.19 v_m$ (\circ) and $1.69 v_m$ ($+$); (v_m is the initial velocity at which the steel balls just reached the potential minimum). The deviation from the straight line must be attributed to irregularities in the surface of the rubber sheet.

Application to a micro-wave triode

After the tests described above the method was applied to the development of a new valve, a micro-wave triode. In such valves the diameter of the grid wire, although small (5μ), is not so very small compared with the grid-to-cathode spacing (fig. 11), in consequence of which the space charge assumes a very tortuous form which cannot be computed accurately.

We proceeded from a more or less plausible distribution of the space charge, computed by assuming the triode divided into sections, parallel to the grid wires and perpendicular to the cathode. These sections were regarded as ideal triodes, lateral deflection being disregarded. The method of computation has been described by Bennett and Peterson¹⁰⁾ amongst others. A number of typical paths

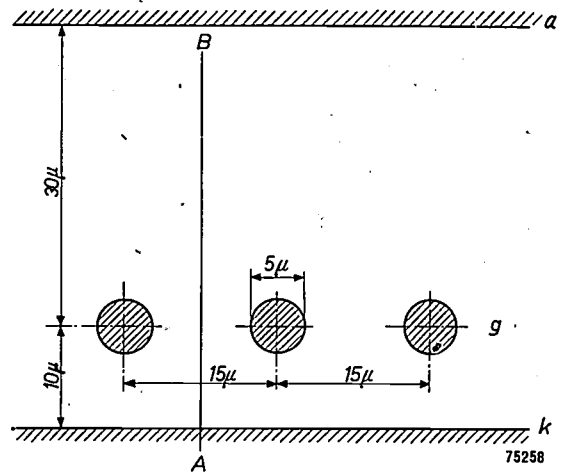


Fig. 11. Electrode configuration in a micro-wave triode. *k* cathode, *g* grid, *a* anode.

in the potential field produced by such a space charge are depicted in fig. 12. The "grid bias" employed corresponds to a high current density,

¹⁰⁾ W. R. Bennett and L. C. Peterson, Bell Syst. tech. J. 28, 303-314, 1949.

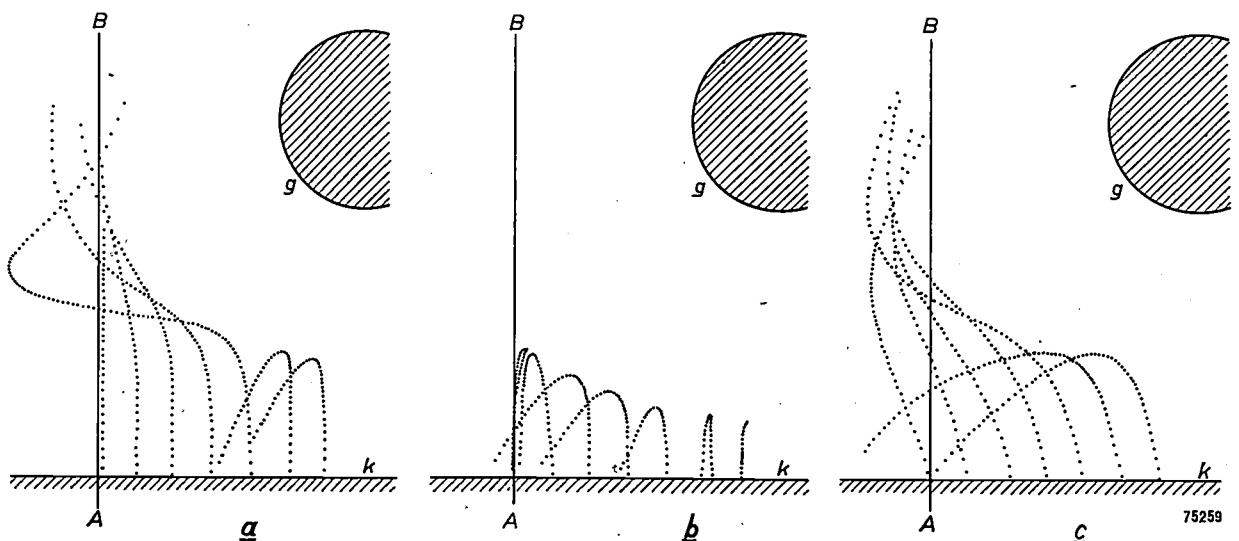


Fig. 12. Electron paths in a micro-wave triode as determined on the rubber sheet. The lines *AB* indicate the centre between two grid wires (cf. fig. 11).
 a) Paths resulting from initial velocity perpendicular to the cathode: electrons passing the grid.
 b) Paths resulting from a lower initial velocity, perpendicular to the cathode: electrons turned back by the grid.
 c) Paths resulting from the same initial velocity as in (a), but started at an angle of 20° from the normal to the cathode.

although sufficiently below saturation to ensure only a low grid current.

The fact that so many electrons are deflected, and to such an extent, by lateral fields, shows that the distribution in the individual sections of the triode cannot be expected to give a fair approximation of the space charge in the entire triode. But, in view of the rapid convergence exhibited in the two preceding examples, it can be said with reasonable certainty that the photographed paths should not be far from the true paths, and that the space charge density computed from the paths as photographed should not differ very greatly from the actual distribution.

The pronounced differences in transit times (number of dots per track) of the electrons should also be noted. This is the reason for the drop in the slope in the higher frequencies.

Appendix: Convergence of the method of successive approximation

In the following we shall give two simple mathematical examples to demonstrate the rate at which the successive approximation process converges.

Both these examples relate to a static, uni-dimensional situation whereby all the electrons move in the positive x direction with the same initial velocity. The zero level of the potential φ is taken such that $\varphi = 0$ when the velocity of the electrons $v = 0$. The electron current, of density j , then gives the space charge density ρ :

$$\rho = \frac{j}{v} = j \sqrt{\frac{m}{2e}} \varphi^{-1/2}, \dots (7)$$

where m = mass and e = charge of the electron.

Poisson's equation (3) assumes the form:

$$\frac{d^2\varphi}{dx^2} = -\frac{\rho}{\epsilon_0} \dots (8)$$

Mathematically, the successive approximation process comes to this, that in (8) we substitute for ρ an arbitrary (simple) function $\rho_0(x)$; the solution $\varphi_0(x)$ of (8) is then substituted in (7) and the function $\rho_1(x)$ thus obtained is once more substituted in (8), the new solution $\varphi_1(x)$ again in (7), and so on.

For the solution of the differential equation (8) the integration constants are each time determined by the given electrode potentials.

In both examples we have also made a direct computation of $\varphi(x)$. By substitution of (7) in (8) a differential equation for φ is obtained which is capable of solution¹¹⁾; the limit to which the approximation method should converge is therefore already known.

The plane diode

The first example refers to a plane diode. We will assume that the electrons leave the cathode ($x = 0$) without initial velocity and that the current is limited by the space charge.

In this case the current density is found from the well-known Child's law (l = anode-cathode distance, φ_a = anode potential):

$$j = -\frac{4}{9} \epsilon_0 \sqrt{\frac{2e}{m}} \frac{\varphi_a^{3/2}}{l^2} \dots (9)$$

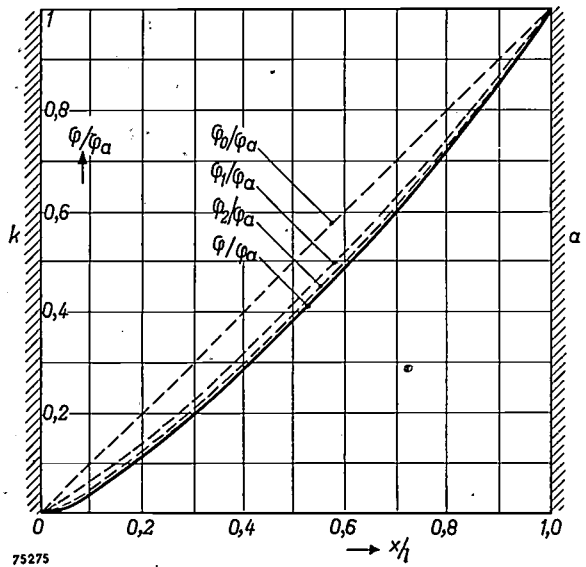


Fig. 13. The method of successive approximation applied to the potential distribution in a plane diode in which the current is limited by the space charge.

With this value of j , together with the boundary conditions $\varphi = 0$ when $x = 0$, and $\varphi = \varphi_a$ when $x = l$, the direct solution of (8) yields Langmuir's law:

$$\frac{\varphi}{\varphi_a} = \left(\frac{x}{l}\right)^{4/3} \dots (10)$$

This potential curve is shown in fig. 13, together with the successive approximations φ_0, φ_1 and φ_2 obtained with $\varphi_0(x) = 0$ as starting point. It will be seen that the second approximation is already almost the exact potential distribution (10).

Tetrode

The second example refers to the space between the screen grid g_2 and the anode of a tetrode, both being assumed to be at the same potential φ_a . The current density j is dependent on the control grid. Direct solution of $\varphi(x)$ is again possible, and yields the potential distribution plotted in fig. 14 which, as

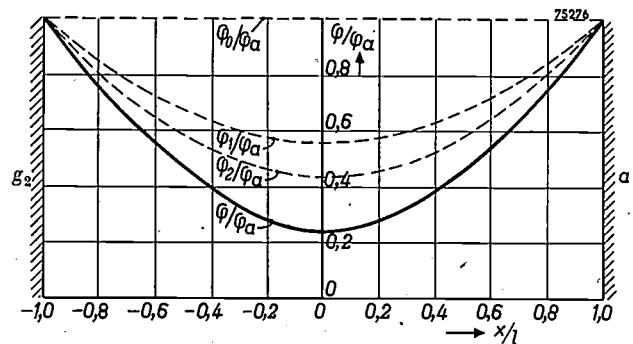


Fig. 14. The method of successive approximation applied to the potential distribution in the space between screen grid g_2 and anode a of a tetrode in which these electrodes are at the same potential φ_a and in which the current density is maximum ($j/j_{max} = 1$). l = half distance g_2 - a .

¹¹⁾ See for example H. Rothe and W. Kleen; Grundlagen und Kennlinien der Elektronenröhren, Akad. Verlagsges. Leipzig 1943, p. 40. The solution as applied to the case of the plane diode is given on p. 20.

might be anticipated, reveals a minimum φ_m in the centre of the space between the electrodes. The depth of this potential minimum increases with j . Moreover, it is found that j cannot be increased beyond a certain maximum¹²⁾, j_{max} , where $\varphi_m/\varphi_a = \frac{1}{4}$. j/j_{max} is employed as parameter, and the curve in fig. 14 is valid for $j/j_{max} = 1$.

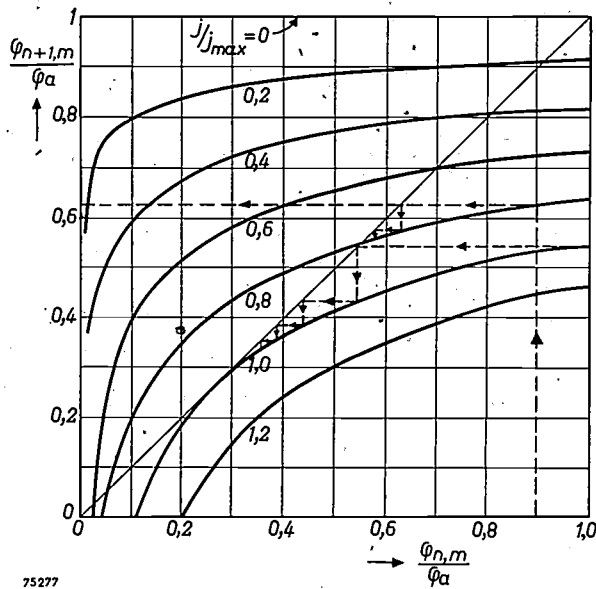


Fig. 15. Graphical representation of the process involved in the method of successive approximation as applied to a tetrode for different values of j/j_{max} .

Apart from the exact solution $\varphi(x)$, fig. 14 includes the successive approximations for φ_0 , φ_1 and φ_2 obtained by starting from $\varrho_0(x) = 0$. It is seen that the second approximation $\varphi_2(x)$ still differs considerably from the true potential distribution. The function $\varphi_1(x)$, as will be seen from a simple calculation, represents a parabola. The function $\varphi_2(x)$ is so complex that the necessary integration to yield $\varphi_3(x)$ is too

¹²⁾ If endeavours be made to increase j beyond j_{max} the potential minimum drops to zero and electrons then commence to return from this minimum, so that the condition $\varrho = j/v$ mentioned in (7) is no longer met. The matter is dealt with fully in the book referred to in footnote¹¹⁾ (p. 40).

difficult to perform; nevertheless, with some sacrifice of accuracy we can proceed, for $\varphi_2(x)$ can be replaced as a close approximation by a parabola having the same minimum $\varphi_{2,m}$. Making use of this parabola we can then obtain $\varphi_3(x)$ and repeat the process. By means of an arbitrary parabola $\varphi_n(x)$ with a minimum $\varphi_{n,m}$, the minimum of the next approximation $\varphi_{n+1,m}$ can be computed as a function of $\varphi_{n,m}$, and this is illustrated in fig. 15, for various values of j/j_{max} .

If we assume for example that $j/j_{max} = 0.8$ and that $\varphi_{0,m}/\varphi_a = 0.9$, we find from the curve for $j/j_{max} = 0.8$ a value of 0.63 for $\varphi_{1,m}/\varphi_a$; this is used as starting point on the abscissa to evaluate $\varphi_{2,m}/\varphi_a$ and so on.

The stepped line between the curve for the selected value of j/j_{max} and the straight line $\varphi_{n+1,m} = \varphi_{n,m}$ demonstrates the rate at which the final result is approached.

The approximation $\varphi_0(x)$ obtained when using $\varrho_0(x) = 0$ as starting point, i.e. the horizontal line in fig. 14, can also be regarded as a parabola with a minimum $\varphi_{0,m} = \varphi_a$. If a start be made with zero space charge, the stepped line in fig. 15 must commence at $\varphi_{0,m}/\varphi_a = 1$.

It will be seen that the convergence is reduced according as j/j_{max} is moved towards unity; if we take $j/j_{max} > 1$, the method of successive approximation is no longer convergent, as already mentioned in footnote¹²⁾.

Summary. The motion of electrons in two-dimensional electric fields can be imitated by rolling steel balls over a surface which is a "model" of the potential field. Such a model can take the form of a uniformly stretched rubber sheet supported in an appropriate manner. It is found that the effect of an electrostatic space charge can be reproduced by applying suitably distributed pressure from beneath the rubber sheet. Since the space charge is caused by the very electrons whose motion it is intended to determine, it is necessary to employ a process of successive approximation.

A description of the apparatus by means of which the pressure is applied is followed by details of four experiments; the first three of these serve as a check on the method and apparatus used, and hence relate to problems the solutions of which have been obtained by other methods (cylindrically shaped rubber sheet, Pierce electron gun, diode with thermal velocity distribution of the electrons).

The fourth experiment concerns the motion of electrons in a micro-wave valve. Two simple mathematical examples (plane diode, tetrode) demonstrating the speed of convergence of the results obtained by successive approximation are given in the appendix.

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 EXPERIMENTAL 100 kW TELEVISION OUTPUT STAGE

by D. ZAAVER.

621.397.645.026.445

A laboratory model of a television output stage has been completed for frequencies up to 68 Mc/s, having a resonance curve suitable for a 6 Mc/s band-width and delivering an output of 100 kW. Output power of this order provides good television reception for some distance beyond the optical horizon.

Television transmitting stations are usually established in the centre, or on the fringe, of densely populated areas. If such an area, as seen from the transmitting aerial, lies largely within the optical horizon, a picture and sound transmitter of a few kilowatts is large enough to cover the whole of it. For a more extensive area, however, considerably more power is needed to give good reception some distance beyond the horizon.

Another reason for the use of powerful transmitters is found in the fact that in many cases the receiving aerials employed are not very effective, either because the owner prefers an indoor aerial to one on the roof, or because the aerial has to be suitable for reception from different directions and at different wavelengths, and this leads to a rather unsatisfactory compromise.

Accordingly, it will be useful to look into the problem of generating considerable power at the frequencies on which television transmitters operate. As far as such frequencies are concerned, the first consideration in any research will be the lowest band of frequencies that has been made available for television: this is from 41 to 68 Mc/s.

A difficulty arises in that the band-width of the video transmitters has to be quite considerable, e.g. 6 Mc/s for a television system of 625 lines and 25 frames per second.

Under this system each TV frame consists of $\frac{4}{3} \times 625^2 \approx 520,000$ picture elements. (The width of the picture is $\frac{4}{3} \times$ the height.) Assuming that the lines are made up of alternate white and black image elements, the time taken to scan two adjacent elements corresponds to one cycle of the video signal. The

whole of the frame is then scanned in $\frac{1}{2} \times 520,000 = 260,000$ cycles. As the image is scanned 25 times per second, the frequency of the video signal is $25 \times 260,000$ c/s = 6.5 Mc/s. A closer study shows that some correction is necessary, in consequence of which the highest modulation frequency used in practice need not exceed 5 Mc/s.

A number of European countries have accepted the 625-line system of the standard set up by the Comité Consultatif International des Radiocommunications (C.C.I.R.)¹. According to this standard the video transmitter should be operated with an asymmetrical side-band, so that the frequency range used shall not be unnecessarily wide, i.e. the transmission should consist of the carrier wave and upper side-band, the lower side-band being for the greater part suppressed. The over-all width of the channel (sound and picture) is then 7 Mc/s.

The ideal frequency curve (aerial current as a function of the frequency) of the complete video transmitter is depicted in *fig. 1* (curve 1); a certain tolerance is necessary, however, and this is represented by the difference between curves 1 and 2, which means that at a frequency of, say 5 Mc/s above the carrier frequency, a drop of 50% = 6 dB is permissible.

The extent to which the frequency curve of a video transmitter is determined by its constituent parts depends on the method of modulation. The modulation systems in use may be grouped into two main classes, viz :

¹ Standards for the international 625-line black and white television system, C.C.I.R., Geneva, 10th October 1950.

1) The R.F. drive voltage at the input of the output stage being kept constant, the grid bias is varied in accordance with the video signal; this is called output stage modulation.

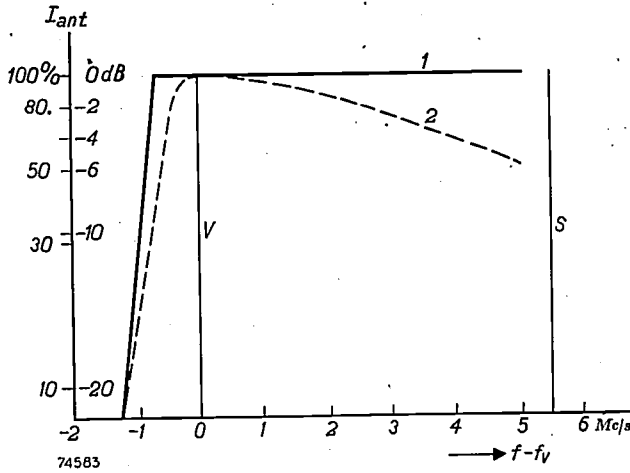


Fig. 1. 1) Ideal, 2) acceptable frequency characteristic of a television transmitter for 625 lines, in accordance with C.C.I.R. standard; aerial current I_{ant} (vertical); difference between the frequency f and the fixed frequency f_v of the video carrier wave (horizontal). The vertical lines V and S represent the video and sound carriers respectively.

2) Modulation in one of the preceding stages, e.g. variation of the grid bias with the video signal in the control stage or one of the stages preceding it: the output stage then receives a modulated R.F. drive voltage and constant bias, and thus operates as a linear amplifier.

Now, the frequency curve of a transmitter with output modulation depends on the modulator, the driver and output stages, the aerial and the filter for suppressing the lower side-band (located between output stage and aerial). With modulation applied to one of the preceding stages, those stages which operate as amplifiers also affect the result.

In view of the object outlined above, viz. to generate considerable output power, this article is mainly concerned with the output stage of the transmitter. This part of the transmitter must be responsible for not more than a mere fraction of the permissible difference between the two curves shown in fig. 1. We have therefore set out to design an output stage tuneable from 48 to 68 Mc/s that will yield a resonance curve which does not drop more than 1 dB at a frequency of 5 Mc/s above the carrier frequency. This output stage is to consist of a single circuit, with resistive load. With a frequency range of 6 Mc/s (5Mc/s for the upper side-band plus 1 Mc/s for the lower), the current flowing in the loading resistor must therefore remain constant to within 1 dB.

That there is to be only one LC circuit means that the conditions are less favourable than in conventional transmitters, in which a transmission line, inductively coupled to the output stage, is used for feeding the aerial.

In order that the transmitter shall radiate as few harmonics as possible (as these may interfere with other channels), it is usual to couple this line inductively to the output circuit; the output and aerial circuits then function as a band-filter which, for the same valve loading, will give a greater resonance width than a single circuit.

This requirement can be met most simply by offsetting the tuning so that the resonance frequency does not coincide with the carrier frequency, but lies slightly higher, e.g. 1.5 Mc/s, as shown for the No. 4 TV channel ²⁾ in fig. 2. All the same, it is anything but simple to obtain a sufficiently flat characteristic over a range of 6 Mc/s, especially when high power is required. The conflicting conditions of a flat resonance curve and high power will now be examined.

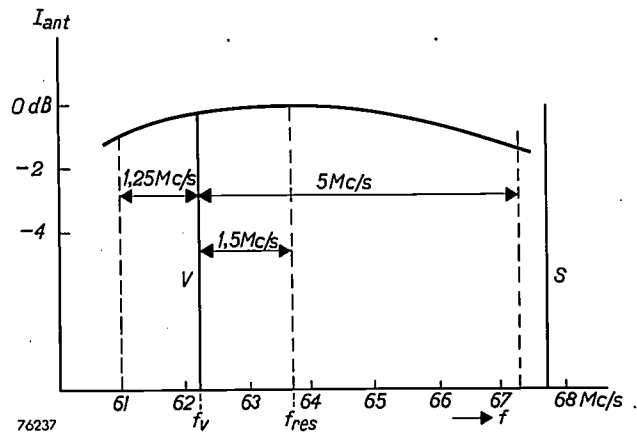


Fig. 2. Resonance curve of the output stage of a video transmitter. f_{res} resonance frequency, f_v frequency of the video carrier wave (TV channel 4). V video carrier, S sound carrier. A satisfactory region of the resonance curve with respect to the frequency band $f_v - 1.25$ Mc/s to $f_v + 5$ Mc/s is ensured by taking f_{res} to be roughly 1.5 Mc/s higher than f_v .

Resonance width and power

Fig. 3a shows the anode circuit of an output stage. For convenience the output circuit is represented by an LC circuit, although in television

²⁾ At an international conference at Stockholm (1952) five plans were prepared for the distribution of wavelengths for European television transmitters in those countries which have accepted the international 625-line standard, and in Great Britain, France, Belgium and Eastern Germany. Some of the channels under the plan for the C.C.I.R. countries are mentioned in this article, hereinafter referred to as channels 2, 3 and 4 (the numbering is unofficial). The video carrier frequency of channel 2 is 48.25 Mc/s, that of channel 3 is 55.25 Mc/s and that of channel 4, 62.25 Mc/s; these all fall within the frequency band of 41 to 68 Mc/s mentioned above.

transmitters working at wavelengths of a few metres, a transmission line is usually employed (as mentioned later). The parallel resistance R represents the effective load of the aerial, as well as the losses in the circuit and the damping produced by the transmitting valves.

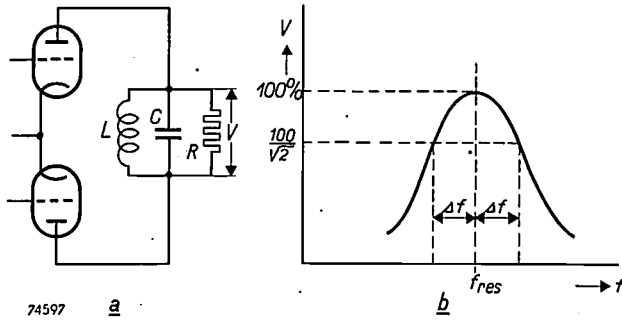


Fig. 3 a) Anode circuit of a push-pull output stage. LC oscillatory circuit with total load R .
 b) Resonance curve: the voltage V across the network LCR as a function of the frequency f with constant alternating grid voltage. The resonance width is $2\Delta f$.

A circuit of this kind gives a resonance curve such as that shown in fig. 3b, the "width" of which is defined as $2\Delta f$, where Δf is the amount by which the frequency must differ from the resonance frequency for the voltage across the circuit to drop to $\frac{1}{\sqrt{2}}$ times the peak value, the grid voltage remaining constant. For Δf we may write:

$$\Delta f = \frac{1}{4\pi CR} \dots \dots \dots (1)$$

Under normal operating conditions (R.F. Class B) the anode current is sinusoidal during one half-cycle (amplitude $I_{a \max}$), and zero during the second half (fig. 4). The amplitude of the first harmonic of this current (broken line) is $\frac{1}{2} I_{a \max}$ and the r.m.s. value is therefore $I_{a \max} / (2\sqrt{2})$. Hence, the power dissipated by the parallel resistance R is:

$$P = \left(\frac{I_{a \max}}{2\sqrt{2}} \right)^2 R,$$

or:

$$R = \frac{8P}{I_{a \max}^2}.$$

Substitution in equation (1) then gives for the half-width of the resonance curve:

$$\Delta f = \frac{I_{a \max}^2}{32\pi CP} \dots \dots \dots (2)$$

From this it is seen that Δf and P are inversely proportional. In order to secure the highest-possible power for a given band-width, the capacitance C

must be as small as possible and the peak current $I_{a \max}$ as high as possible. This peak current is limited by the saturation current of the cathodes. For the object in view, that is high power with large band-width, the valves used should therefore have a high saturation current, and this is obtained without entailing excessively high heater power by using thoriated tungsten cathodes; as explained in a recent issue of this Review³⁾, this material can now also be used for high-power valves, owing to the introduction of more effective getters.

In the design of output stages the question arises whether triodes or tetrodes should be used, and one of the points to be considered in this choice is the effect of the internal resistance R_i . Let us take two examples, on referring to tetrodes and the other to triodes, for the same maximum anode dissipation. The output capacitance of these valves is roughly the same, and determines to a large extent the value of the capacitance C in equation (1). The resistance R will also be assumed to be the same for both types so that, in accordance with (1), the two circuits will have similar resonance widths. The resistance R may be regarded as consisting of three resistances in parallel, viz. resistance R_a representing the effective load of the aerial, $R_{L,C}$ corresponding to the losses in the oscillatory circuit, and another resistance to represent the damping due to the valves; under Class B conditions the last-mentioned is $2R_i$. In tetrodes R_i is so high that its effect can be disregarded. In triodes, however, the value of R_i is such that R_a must be appreciably higher than in tetrodes in order to sustain the assumption of equal total resistances R (the loss resistances $R_{L,C}$ are assumed equal). The higher value of R_a is an advantage, for the required bandwidth was obtained by choosing R_a lower than the value that matches the valve impedance; with the higher value of R_a now required the matching is improved, i.e. the output is increased.

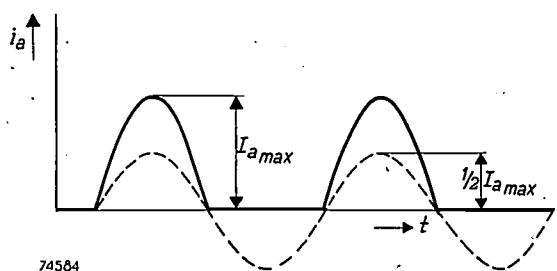


Fig. 4. Full line: anode current i_a of one of the valves in an R.F. Class B push-pull circuit, as a function of the time t . Dotted line: the first harmonic of i_a .

This advantage of the triode is offset by two drawbacks; firstly, the value of R_i varies in consequence of the modulation; this is referred to again at the end of the article. Secondly, triodes in general require more drive power than tetrodes of the same rating.

In practice both kinds of valve give satisfaction; tetrodes are used at the transmitter at Lopik (Netherlands). The tests described in the following all relate to triodes, however.

³⁾ E. G. Dorgelo, Philips tech. Rev. 14, 226-234, 1953 (No. 8).

In order to ascertain whether a resonance curve such as that shown in fig. 2 can be obtained simultaneously with high output power, and also whether the modulation characteristic will then be satisfactory, we have constructed an experimental unit comprising an output stage with driver stage and preceding circuits, feeding a load resistance (dummy aerial). The frequency is variable from 48 to 68 Mc/s, which range covers television channels 2, 3 and 4.

We shall first say something about the output circuit, then pass on to the different sections and finally describe the results obtained.

Circuit of the output stage.

An advantage of push-pull output circuits is that a point of symmetry occurs at the input of the circuit, at which there is no R.F. voltage with respect to earth. When modulation is applied to the output stage it can be introduced directly at this point, thus keeping the load on the modulator at a minimum. On the other hand, asymmetrical circuits (a single valve, or several in parallel) entail, in the range of frequencies with which we are concerned, the use of by-passing capacitances (coaxial systems), which constitute an extra load.

At the same time, it should be added that this capacitance can be avoided very simply by increasing the length of the coaxial system by one quarter-wavelength. However, this is practicable only when the wavelength is short enough (frequencies above 100 Mc/s); at such frequencies an asymmetrical circuit is therefore often employed.

It is necessary next to consider whether the valves are to be operated in an earthed-cathode, or earthed-grid circuit, i.e. whether the cathodes or the grids are to form the common point in the input and output circuits.

A conventional circuit for short-wave transmitters with small band-width is the earthed-cathode circuit shown in fig. 5a. To avoid coupling between the anode and grid circuit via the anode-to-grid capacitance, such that the stage oscillates, the grid of each valve is connected to the anode of the other through a neutralising capacitor C_n . As long as the frequency is comparatively low, coupling is completely eliminated when $C_n = C_{ag}$.

In an earthed-cathode circuit neutralised in this way there are two capacitive branches in parallel with the coil in the anode circuit, each consisting of C_{ag} and C_n in series. Since $C_n = C_{ag}$, these branches function as a single capacitance (C_{ag}). In the earthed-grid circuit, on the other

hand (fig. 5b), we are mainly concerned with a branch in parallel with the anode coil, comprising the two capacitances C_{ag} in series, which behave as a capacitance $\frac{1}{2} C_{ag}$ (In both cases the slight effect of the capacitance C_{af} between anode and cathode may be ignored; C_{af} is roughly $\frac{1}{25} C_{ag}$.) With the earthed-grid circuit the contribution of the valves to the over-all output capacitance is thus roughly half that of the earthed-cathode circuit. This is of considerable importance where large transmitting valves are employed, as $\frac{1}{2} C_{ag}$ (or C_{ag}) then constitutes the greater part of the total output capacitance. A low capacitance is an advantage firstly from the point of view of the resonance width and output power rating (see Eq. (2)) and secondly, because the alternating current in the LC circuit is lower and the losses smaller for given frequency and amplitude of the R.F. voltage.

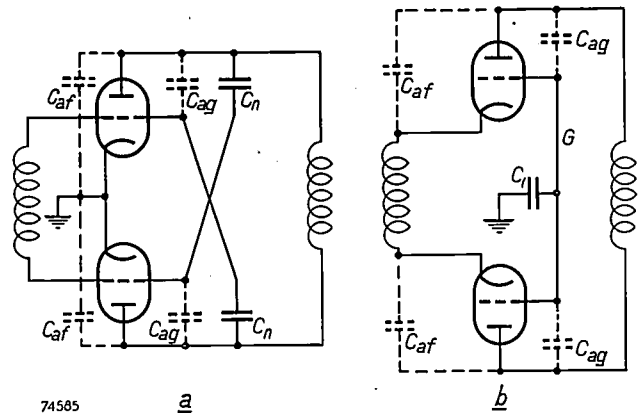


Fig. 5. Push-pull stage with a) earthed cathodes, b) earthed grids. C_{ag} anode-grid capacitance, C_{af} anode-cathode capacitance, C_n neutralising capacitor, C_1 capacitance between earth and a screen plate G used for connecting the grids.

In the neutralised, earthed-cathode circuit (fig. 5a) the self-inductance of the leads to the neutralising capacitor, at very high frequencies, affects the value that C_n should assume. The precise value of C_n thus becomes dependent on frequency to such an extent that insufficient neutralising is obtained throughout the frequency band of a videotransmitter. This applies equally to any other neutrodyne circuit. In this respect, too, the earthed-grid circuit is much better than the earthed-cathode arrangement, provided that the valves are such that the grid serves as an effective screen between the input and output circuits; such screening should be not only electrical, but also magnetic. At higher frequencies the current flows only in the thin outer layer of the conductor (skin effect), and this effect can be put to good use in that, if matters are carefully arranged, the input and output circuits can be so separated that neutralising capacitors are no longer necessary.

Fig. 6 illustrates a suitable arrangement. The ring-seal grid contacts of the valves are mounted in a common metal screen plate. Owing to the skin effect, the

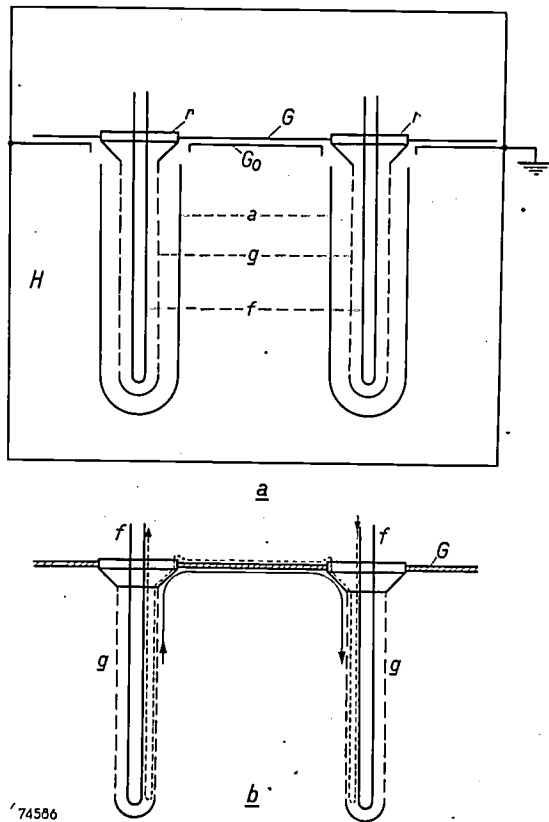


Fig. 6. a) Cross-sectional diagram of two transmitting valves with earthed grids. a anodes, g grids, f filaments, r ring-shaped grid seals connected to common screen plate G. The plate G_0 , which is part of the chassis H, forms with G a capacitor (C_1 in fig. 5b). b) Path of the R.F. currents through filaments, screen plate and grids. The input current flows in the skin shown by the dotted arrow, and the output current in the skin indicated by the full arrow. The only residual coupling between the circuits is that which occurs by way of the mesh of the grids.

input current — as far as the grids and screening plate are concerned — flows only on the upper face of the plate and the inner faces of the grids, and the output current only on the underside of the plate and the outer faces of the grids. The two currents are thus separated by the body of the screen plate and grids.

Imperfections in the screening occur only as a result of the presence of the grid meshes; these give rise to a certain capacitance C_{af} between anode and cathode, which is accompanied by mutual inductance between the circuits. Below a certain frequency the coupling across C_{af} is dominant; above that frequency, the coupling through the mutual inductance becomes the more important. This is shown by the so-called "interaction" test in which an RF voltage of variable frequency and constant

amplitude is applied between the cold cathodes. The RF voltage in the output circuit resulting from the imperfect screening is then measured at the output as a function of the frequency. The result of such a test, as applied to the output circuit of our experimental equipment arranged as shown in fig. 6, is demonstrated in fig 7. A minimum occurs at 62 Mc/s; at that point the effect of C_{af} counterbalances that of the mutual inductance. (The minimum is not at zero as a result of losses.) Throughout the whole frequency range from 48 to 68 Mc/s (channels 2, 3 and 4), this interaction is so slight that the stage under discussion is stable in operation, without taking further measures to ensure this.

As the grids have to be biased, they cannot be connected directly to the chassis, but only via a capacitor. As will be seen from fig. 6, this capacitor is formed by the screening plate and a second metal plate which is actually part of the chassis itself; these plates are separated by a solid dielectric such as mica or polythene.

Although our experimental unit is not adapted for modulation, a few remarks on this subject will not be out of place.

The arrangement shown in fig. 6 is suitable for use as an amplifier, the modulation being therefore applied to a preceding stage; the modulated R.F. drive voltage is applied between the cathodes of the output valves. For modulation in the output stage itself, the video signal, in this circuit, is applied between the cathodes and earth, but this has the disadvantage that the modulator is very heavily loaded, not only

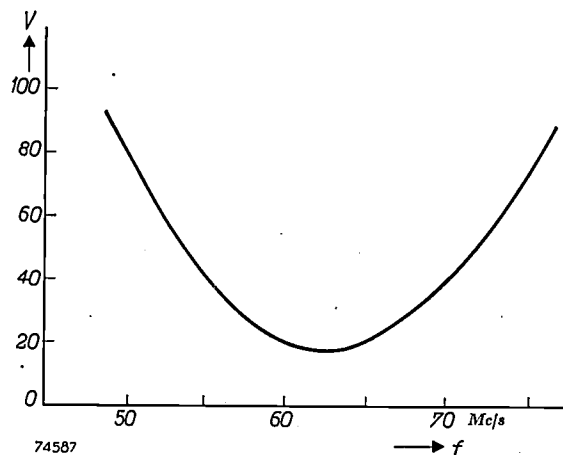


Fig. 7. Interaction: measured voltage V between the anodes of the output stage (on relative scale) as a function of the frequency f , with constant input voltage and cold cathodes.

by the valve capacitances, but also by the capacitance of the filament transformer. Preference is therefore sometimes given to the circuit shown in fig. 8a in which the grids are not connected directly to each other, but with two capacitors in series between them (for symmetry), and a choke with a centre tap to which the video signal is applied. Fig. 8b shows a practical form of the circuit.

Owing to the absence of a continuous screening plate, the paths of the current in the input and output circuits are not fully separated, and a certain amount of coupling occurs. This can be represented, together with the magnetic coupling due to the meshes of the grid, by a self-inductance L' in series with

delivered is $I_a(V_i + V_a)$; of this, apart from the losses already mentioned, the quantity $I_a V_i$ must be supplied by the driver stage, which thus has to be larger than for an earthed-cathode output

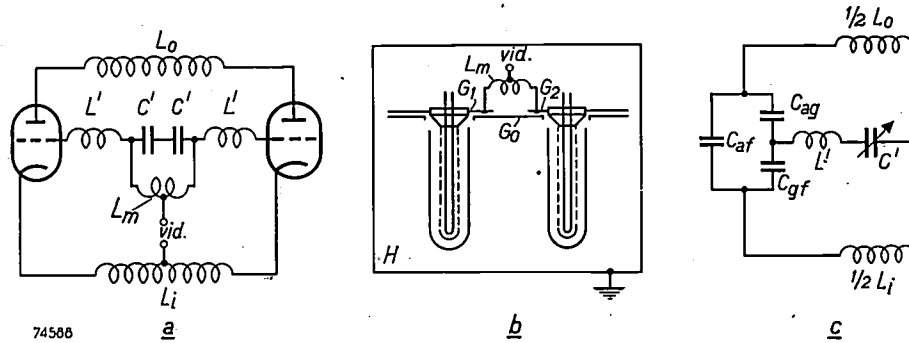


Fig. 8. a) Circuit for modulation in the output stage. The grids of the valves are connected to each other capacitively ($C'-C'$) and the video signal *vid* is applied through a centre-tapped choke (L_m). L_i input circuit, L_o output circuit. L' stray inductance. b) Practical form of circuit a). The capacitors C' are formed by screening plates G_1 and G_2 on the one hand and plate G_0 (not connected to the chassis H) on the other. c) Equivalent circuit of one half of the arrangement. C_{ag} , C_{gf} and C_{af} valve capacitances; L' stray inductance. Coupling between circuits can be reduced to zero by a suitable value of C' .

each grid (fig. 8a). The equivalent network for one half of the circuit is depicted in fig. 8c, in which the valve capacitances (C_{ag} , C_{gf} and C_{af}) are also shown. It can be shown that the coupling between the input and output circuits is zero when

$$X' = \frac{-X_{fg} X_{ag}}{X_{fg} + X_{ag} + X_{af}}$$

where $X' = \omega L' - \frac{1}{\omega C'}$, $X_{fg} = \frac{1}{\omega C_{fg}}$, etc. (with $\omega = 2\pi f$).

This equation is satisfied by giving C' a certain value. This "neutrodyning" is again dependent on frequency, to such a degree that stabilization is achieved only in a small range of frequencies. In Great Britain, where the 405-line system is employed and the frequency band is accordingly fairly narrow, the latter method has been adopted, but in transmitters operating with 625 lines or more it is less suitable owing to the high "selectivity" of the neutrodyne.

When comparing the merits of earthed-cathode and earthed-grid circuits we must also take into consideration the driver stage. The driver stage of an earthed-cathode circuit delivers power which serves only to cover the grid dissipation in the output valves and grid resistors, together with the losses in the input circuit. On the other hand, in the earthed-grid circuit the driver stage must deliver a certain amount of power which is passed on to the load in the output stage. This is illustrated in fig. 9, which shows the simplest form of earthed-grid circuit. The A.C. voltage across the load resistance R is the sum of the alternating input voltage V_i and the alternating voltage across the valve V_a . If we denote the anode alternating current by I_a and if I_a , V_i and V_a be regarded as r.m.s. values, the power

stage. This disadvantage, however, does not outweigh the above-mentioned advantage of the lower output capacitance and greater stability of the earthed-grid circuit.

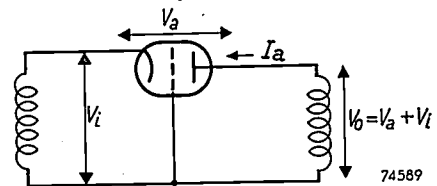


Fig. 9. Amplification stage with earthed grid. The alternating output voltage V_o is the sum of the alternating input voltage V_i and the alternating voltage V_a between anode and cathode; the power applied to the input is thus augmented at the output (apart from losses).

Transmission lines

Oscillatory circuits for metre wavelengths are best made in the form of transmission lines, either as parallel line or coaxial conductors. Both types are employed in the present equipment. The parallel lines may be seen in the diagram of the output stage in fig. 10, in which G is the screening plate mentioned above, to which the grids are connected. The cathodes are connected to one end of a two-wire line (Le_i) to form the input, the other end being shorted by a bridge piece B_1 . Because of the relatively high input capacitance of the valves, the filaments themselves constitute a not unimportant part of the input circuit, and the external conductors must therefore be very short. The driver stage (S) is connected to a suitable point on these

conductors, which has to be sufficiently far removed from the shorting piece. To prevent the tapping point from falling too near the filament connections (and thus being inaccessible), the input circuit is extended by one half-wavelength, to about $\frac{3}{4}$ wavelength. Tuning is effected by moving the shorting bridge.

self-inductance⁴), and adjustment of the shorting bridge varies the self-inductance and therefore also the natural frequency of the circuit. As we have already shown above, the effective capacitance in the circuit is, in the main, $\frac{1}{2} C_{ag}$, since the two anode-to-grid capacitances are in series. (Strictly speaking, C_{ag} is not concentrated at any one point,

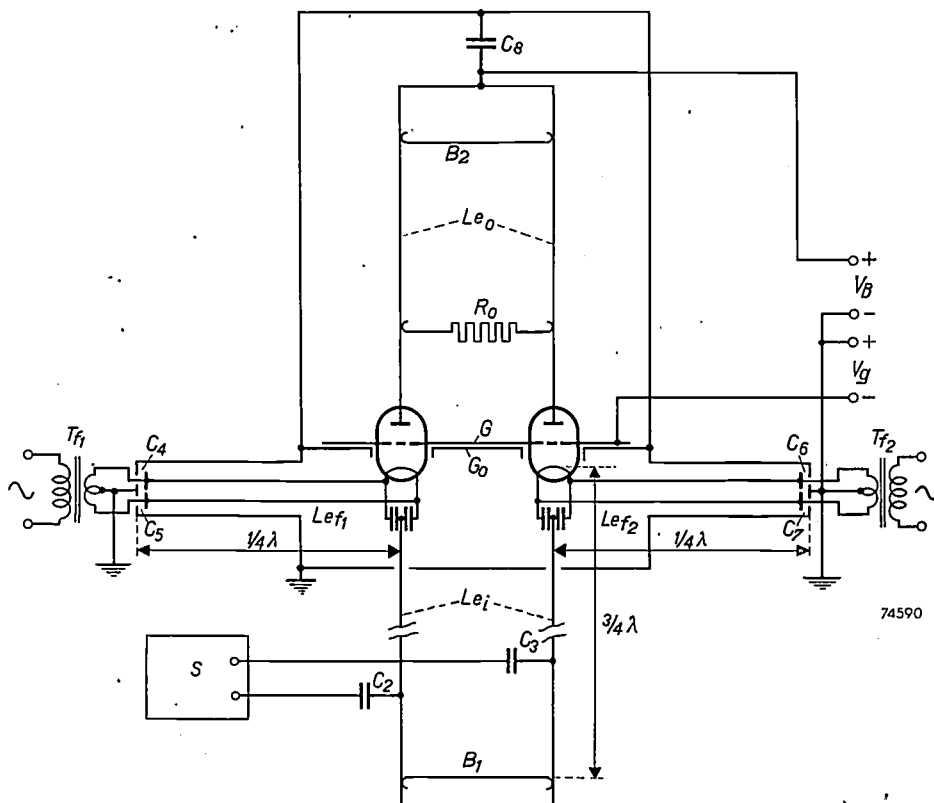


Fig. 10. Circuit diagram of the output stage. *S* driver stage coupled through capacitors C_2 and C_3 to the input transmission line Le_i with tuning bridge B_1 . Le_{f_1}, Le_{f_2} transmission lines through which transformers T_{f_1} and T_{f_2} feed the filaments. C_4-C_5 and C_6-C_7 capacitive tuning bridges for these transmission lines. G screen plate. G_0 earthed plate (Cf fig. 6). Le_o output transmission line with tuner B_2 and movable loading resistor R_0 . C_8 by-pass capacitor. V_g D.C. grid voltage. V_B H.T. supply.

The filament transformers (T_{f_1}, T_{f_2}) are connected by a quarter-wavelength coaxial cable (Le_{f_1}, Le_{f_2}). At one quarter-wavelength from the filament, the filament leads are connected capacitively to the outer conductor (capacitances C_4-C_5 and C_6-C_7). The secondary winding of the filament transformer is earthed.

The anode or output circuit constitutes a fourth transmission line, but in this case the arrangement is more complicated, since the valves are an integral part of the system. Each of the transmitting valves is contained in a cylinder (*C*, fig. 11a). The two cylinders together function as a short two-wire line, shorted by a movable bridge (B_2 in figs. 10 and 11a). As is known, a line of this kind behaves like a

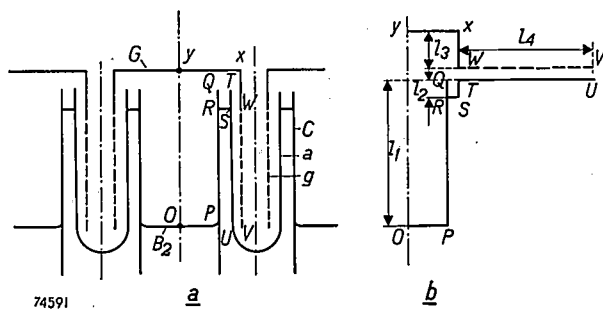


Fig. 11.a) Output transmission line. g grid, a anode, C metal cylinder enclosing anode, with movable tuning bridge B_2 , G screen plate.
 b) Arrangement of the two-wire and coaxial lines showing lengths l_1, l_2 and l_3, l_4 . The points $O, P, Q, \dots X, Y$ correspond to those in (a).

⁴) For this and other properties of transmission lines see article by C. G. A. von Lindern and G. de Vries, Philips tech. Rev. 6, 240-249, 1941.

but is the equivalent value of the open coaxial line formed by the anode and grid). Besides this, there is the capacitance of the two-wire line represented by the two cylinders, but as this line is short compared with the wavelength λ (roughly $1/15 \lambda$), the capacitance is negligibly small.

The transmission line formed by the anode circuit is shown diagrammatically in fig. 11b. Here are seen a shorted two-wire line (the two cylinders and the shorting plate) of which the length is l_1 , two shorted coaxial lines (l_2 and l_3) and an open line, shown in the figure as a two-wire line, representing the grid and anode, of length l_4 . According to equations (19) and (24) in the article referred to in footnote ⁴) the input impedance Z_k of a shorted line — provided length l is small compared with the wavelength (as in this case) — may be expressed as:

$$Z_k = j \zeta \tan 2\pi \frac{l}{\lambda}$$

and the input impedance Z_0 of an open line as:

$$Z_0 = -j \zeta \cot 2\pi \frac{l}{\lambda},$$

where ζ is the characteristic impedance of the line, a real and positive quantity which is a known function of the dimensions. From these formulae it is seen that the shorted line does indeed function as a self-inductance, and the open system as a capacitance.

The natural frequencies of an electrical system are those at which current can flow in the absence of an electromotive force, that is, when the sum of all the reactances is zero. Applying this to the system depicted in fig. 11b, we find that:

$$j \zeta_1 \tan 2\pi \frac{l_1}{\lambda_{res}} + j \zeta_2 \tan 2\pi \frac{l_2}{\lambda_{res}} + j \zeta_3 \tan 2\pi \frac{l_3}{\lambda_{res}} - j \zeta_4 \cot 2\pi \frac{l_4}{\lambda_{res}} = 0.$$

This expression can be employed for estimating the natural frequency λ_{res} for given dimensions.

The curve shown in fig. 12 shows how the measured resonance frequency of the anode circuit

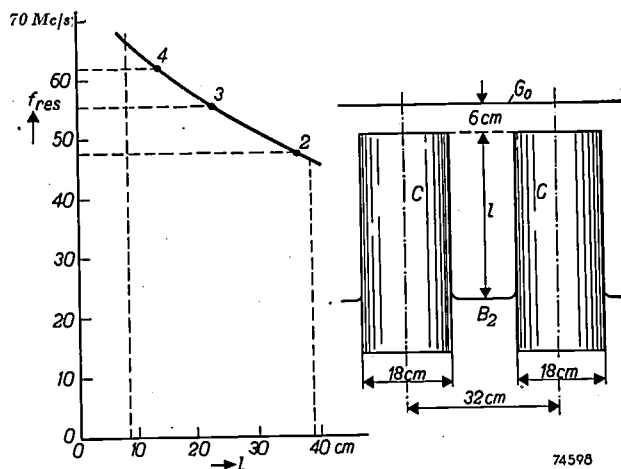


Fig. 12. Resonance frequency f_{res} of the output transmission line plotted against distance l of the tuner B_2 from the top edge of the cylinders C (see fig. on right). The points 2, 3 and 4 correspond to the carrier frequencies of TV channels 2, 3 and 4.

depends on the position of the shorting bridge; it will be seen that the system can be tuned to TV channels 2, 3 and 4, the video carriers of which have frequencies of 48.25, 55.25 and 62.25 Mc/s respectively. As mentioned above (fig. 2), it is an advantage to choose the resonance frequency of the anode circuit somewhat higher.

Constructional details

The output stage

Two type TBW 12/100 valves are used in the output stage. This valve was described fully in a recent issue of this Review (see article referred to in footnote ³). The arrangement of the grid is such that the valve is suitable for earthed-grid circuits; the cathode is of thoriated tungsten and the anode is water-cooled. This valve was designed for operating at frequencies up to 30 Mc/s, but in our test it was used successfully at frequencies up to about 70 Mc/s.

Fig. 13 depicts the actual output stage and shows the two cylinders in the output circuit, with tuning bridge.

A cross-section of one of the cylinders, with valve in position, is shown in fig. 14. That part of the grid which lies above the anode seal is in effect a cylinder with no apertures in the wall; the anode is extended some distance above the seal, which prevents the glass from being struck by the electrons. Cooling air is directed towards the anode and grid-seal from a perforated ring R which is standard with the valve TBW 12/100. The coaxial cylinder is extended right up to this ring, which results in the following advantages:

- 1) Since the potential is the same at all parts of the cylinder at the same height, the anode seal is then in a field-free space and no dielectric losses or consequent heating occurs in the glass.
- 2) The greater length of the cylinders facilitates adjustment of the loading resistor (see below).
- 3) Proximity effect at the anode seal is reduced.

The last-mentioned point can be elucidated as follows. As is known, skin effect occurs at high frequencies, this being manifested as an increase in resistance; it is actually a result of non-uniform current distribution due to the magnetic field of the current itself. Proximity effect occurs between two or more conductors, because the distribution of the current in the one is influenced by the magnetic field of the other(s) ⁵). The current density is thus

⁵) See for example Philips tech. Rev. 6, 220, 1941, or F. E. Terman, Radio Engineers' Handbook, McGraw Hill, New York, 1943, pp. 36-37.

increased at those sides of the conductors which face each other, and this results in a further increase in the resistance. When two valves are mounted close to each other without the screening effect of the cylinders already mentioned, the high-frequency anode current would not be uniformly distributed round the periphery of the anodes, but would be highly concentrated at the sides which face each other, and this would entail some risk of overheating at the anode seals. The above-mentioned extension of the cylinders ensures that the current distribution is much more uniform. Proximity effect does occur in the cylinders themselves, but this can do no harm.

Not only the tuning bridge, but also the load resistor (R_0 in figs. 10 and 13) can be moved along the cylinders. The voltage between points at equal height on the cylinders increases from the bottom upwards, so that the resistance forms a greater load according as it is pushed upwards; this provides a continuously variable load with constant resistance.

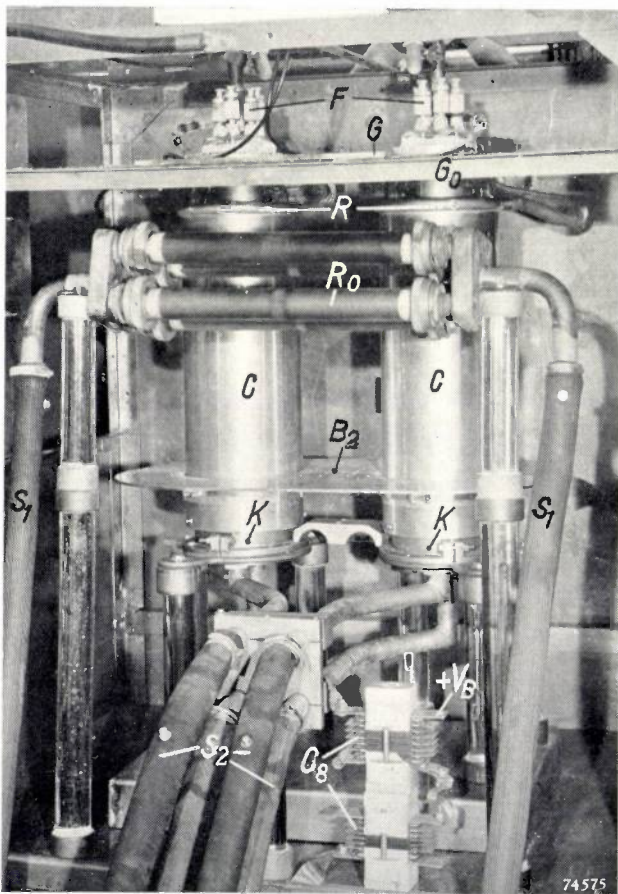


Fig. 13. Photograph of the output stage. *F* filament connections of the valves (TBW 12/100). *G* screen plate connecting the grids and insulated from plate G_0 which is in contact with the chassis (see fig. 10). *R* perforated rings through which cooling air is forced against the glass part of the transmitting valves. *C* cylinders of the output transmission system, with tuner B_2 and loading resistor R_0 , both movable. *K* cooler; S_1 , S_2 cooling-water hoses for the loading resistor and valves. $+V_B$ supply voltage terminal. C_8 by-pass capacitor.

The load resistor itself consists of four ceramic tubes carrying a layer of sintered carbon, and through which cooling-water is run. With a flow of water of 25 litres/min, 25 kW per resistor tube can be dissipated. The four resistors are connected in parallel and, as their length is about 0.1λ , the resistance value (120Ω) is practically independent of frequency.

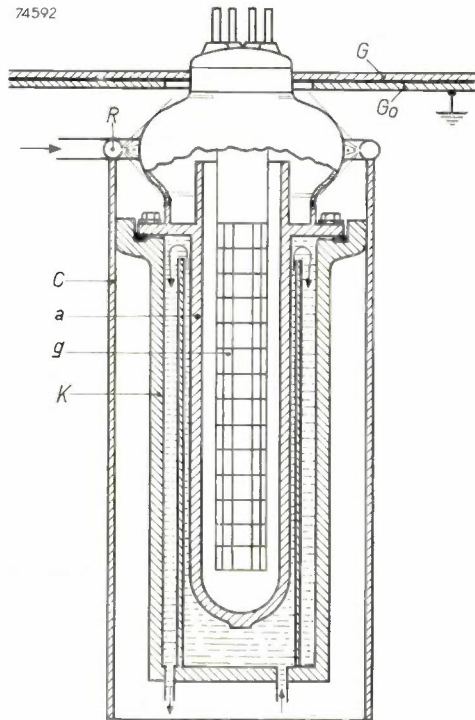


Fig. 14. Cross-section through one of the valves (TBW 12/100) in the output stage. *a* anode, *g* grid. For meanings of the other letters see fig. 13.

The oscillator and other stages

In view of the required constancy of the carrier frequency in TV transmitters, this frequency is invariably generated by a crystal controlled oscillator. Between the oscillator (which does not deliver very much power) and the output stage, a number of stages of R.F. amplification are required. In our experimental equipment, however, a variable frequency was needed, for which reason the stage for driving the output stage has been made up as a variable-frequency driver. The amount of power required for driving an output stage delivering 100 kW is 15 to 20 kW, and it is not so simple to construct an oscillator to give this amount of power, without parasitic oscillation, and giving a reasonably constant output voltage within the whole range from 48 to 68 Mc/s. The arrangement mentioned above, consisting of an oscillator of low power with subsequent stages of amplification, is by far the better of the two. This method was therefore

adopted, the frequency being of course variable; see block diagram in *fig. 15*.

The oscillator (*O*) has two type TB 2.5/300 valves⁶⁾ in push-pull. To vary the frequency either the capacitance or the self-inductance can be varied. An *LC* circuit with variable *C* is incorporated

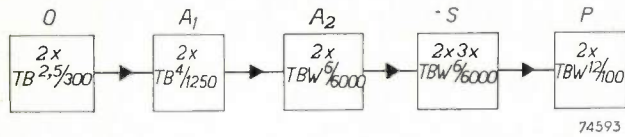


Fig. 15. Block diagram of the experimental equipment. *O* oscillator with variable frequency. *A*₁, *A*₂ amplifiers. *S* control stage, *P* output stage. The numbers of valves and the valve types are shown in the blocks.

between the grids, with another *LC* circuit, of which *L* is variable (variometer) between the anodes. The output voltage thus obtained is sufficiently constant throughout the whole range of frequencies. The oscillator is provided with a frequency-calibrated scale.

Between the oscillator and the driver stage two R.F. amplifiers or sub-driver stages are also ne-

cessary (*A*₁ and *A*₂, *fig. 15*). These are operated with two TB 4/1250 and two TBW 6/6000 valves in earthed-grid push-pull circuits.

The driver stage proper (*S*, *fig. 15*) is also in push-pull with earthed grids, using three type TBW 6/6000 valves on each side (*fig. 16*). The cooling-water inlet and outlet system consists of concentric tubes, the outer one of which simultaneously serves as the transmission line from which the output stage is driven. As in the output stage, use is made of a screening plate connected to the grids and earthed capacitively. The only reason why each valve has its own screening plate (as will be seen from *fig. 16*), is that it is thus possible to measure the grid current of each valve separately.

Measuring facilities

Measurement of the output power as a function of the frequency must be carried out with constant input voltage at the output stage. In effect, then, we measure the input voltage and the output power. As already mentioned, the frequency is read from the scale on the oscillator.

Measurement of the input voltage

The diagram in *fig. 17a* illustrates the method of measuring the input voltage. By reason of the capacitance between one of the input lines (*Le_i*) and a copper probe plate (*D*) mounted in the vicinity, RF voltage occurs on this plate. This voltage is passed to a germanium diode (*Ge*), a smoothing resistor (*R*) and a capacitor (*C*), and is measured on a relative scale by means of a D.C. instrument (microammeter).

The difficulty of a measurement of this kind is one of avoiding the interference from the strong magnetic and electric alternating fields in the region of the output stage, and the method employed to prevent such interference will be seen in *fig. 17b*, which shows a diagram of the measuring equipment. The probe plate is located in or near the open end of a screening tube which contains the germanium diode, the resistor and the capacitor; the plate can be moved away from, or drawn into the tube, to suit the required sensitivity. A flexible coaxial cable connects this unit to the microammeter, which is placed in a metal box to screen it from R.F. currents which might otherwise damage it.

Another point for consideration is the frequency-dependence of the measurements. When the fre-

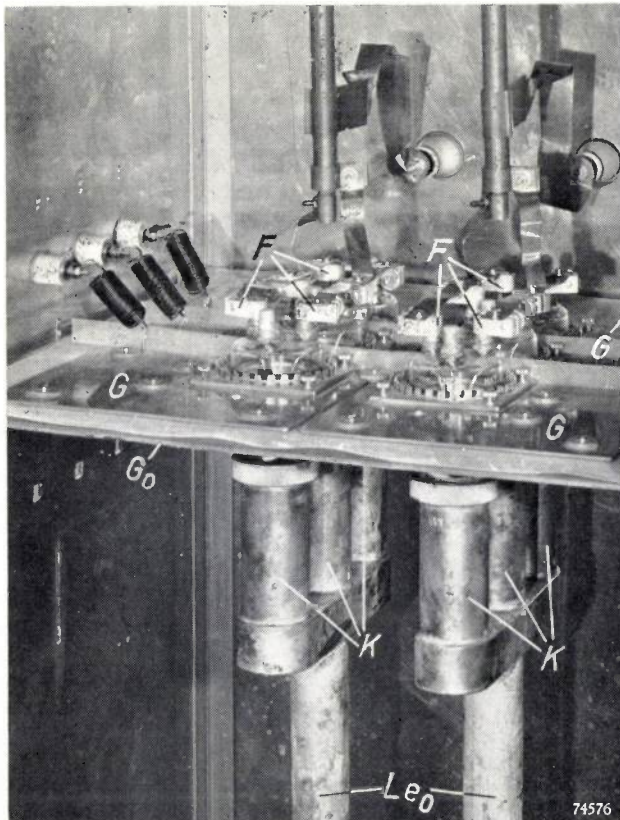


Fig. 16. Interior view of the driver stage with three type TBW 6/6000 valves on each side of an earthed-grid push-pull circuit. *F* filament connections. *G* screening plate. *G*₀ plate in contact with panel. *K* coolers. *Le*₀ output transmission line.

⁶⁾ A description of the transmitting triode TB 2.5/300 is given in Philips tech. Rev. 10, 273-281, 1949.

⁷⁾ The TB 4/1250 is a valve similar to the TB 2.5/300, but has a higher power rating; the TBW 6/6000 is similar to the TBW 12/100, but is of lower power rating.

quency is varied the potential maxima and minima travel along the transmission line and the deflection of the meter changes. Now, the voltage in the direction of the transmission line varies considerably in the region of a minimum, but only slightly in the vicinity of a maximum, and it is therefore an advantage to mount the probe plate at a point where a maximum occurs. If this is done, the error due to displacement of the maximum, within the frequency range of a single channel, can be disregarded. For other channels the position of the relevant maximum has first to be located.

Measuring the output power

With constant water flow the increase in temperature of the cooling-water in the load resistor R_0 is a measure of the power P_0 dissipated. The temperature of the inlet water being constant, the thermometer used for measuring the temperature at the water outlet may be calibrated in kW to give a direct reading of P_0 .

Results obtained with the experimental equipment

It is found that the equipment is capable of delivering a continuous output of 100 kW. The suitability of such equipment, when complete with modulator, aerial, etc., as a television transmitter for 100 kW peak output, can be confirmed only when it is known that the bandwidth is sufficient and the modulation curve sufficiently linear. Results concerning these details are given below.

As a basis we have taken the C.C.I.R. standard, case which specifies negative modulation⁸⁾. In this the aerial current is at its maximum (100%) for the peak of the synchronisation signals; it drops to 75% on black level (lowest image brightness) and to 10% on white level (highest image brightness). Hence, at black level, the transmission power is $0.75^2 = 56\%$ of the peak power. As the TBW 12/100 valves can deliver 100 kW continuously, there is therefore a very wide margin of safety with 100 kW at the synchronisation peaks.

Resonance curves

Measurements were taken of the temperature increase $\Delta\vartheta$ of the cooling-water in the load resistor as a function of the frequency, in the region of the resonance frequency; this was effected with constant tuning of the anode circuit and constant input voltage, for different values of the load, which was varied by changing the position of the resistor R_0 .

Figs. 18a and b show the "aerial current" $I_{ant}(\sim\sqrt{\Delta\vartheta})$ plotted against the frequency for TV channels 2 and 4. At resonance, I_{ant} was found to be 75% of the value relevant to 100 kW; hence these figures correspond to black level in a transmitter with 100 kW peak output. It appears that within a frequency range of 6 Mc/s, I_{ant} is less than 1 dB below the resonance value, thus meeting the requirement originally laid down.

As pointed out in the foregoing, this requirement is rather on the high side, but we consider that it should be, for the following reason. With modulation, the anode direct current is

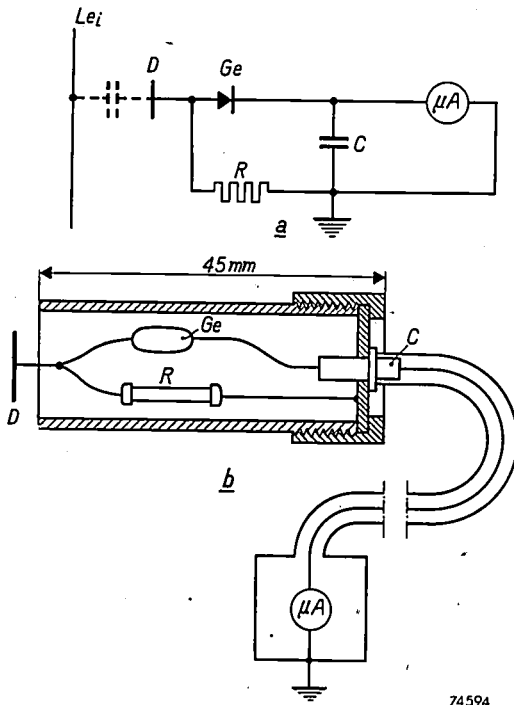


Fig. 17. a) Measuring circuit for the alternating input voltage V_i . Le_i one of the conductors of the input transmission line, D probe plate. Ge germanium diode. R and C smoothing resistor and capacitor. μA microammeter. b) The assembled unit. Components D , Ge , R and C are housed in an earthed metal tube for screening purposes. The meter is connected by means of coaxial cable and is also screened.

As pointed out, the measurement is only relative, but this is sufficient if its sole purpose is to ensure that the input voltage remains constant. With a little extra trouble the equipment could be calibrated so that the microammeter gives direct indication of the R.F. voltage between cathode and grid. In so doing we would be making use of the fact that grid current commences to flow when the peak R.F. voltage between cathode and grid is practically equal to the D.C. voltage between these electrodes. This D.C. voltage (V_g in fig. 10) is variable and can be directly measured. The occurrence of grid current can be observed by means of a D.C. instrument in series with the voltage source V_g .

⁸⁾ See article mentioned in footnote 1), or Philips tech. Rev. 13, 313, 1952 (No. 11).

lower than the value relating to black level, and this is accompanied by an increase in the internal resistance R_i of the valves. The valve damping thus drops and the bandwidth is accordingly reduced. Hence, when the modulation depth is increased, the bandwidth decreases.

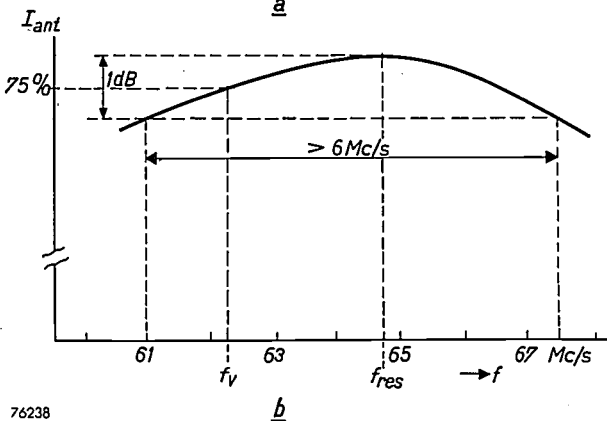
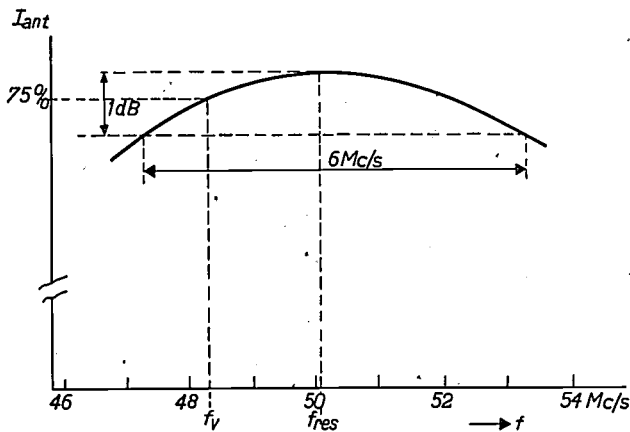


Fig. 18. Resonance curve $I_{ant} = f(f)$ of the output stage tuned to a) TV channel 2, b) channel 4. Within a frequency range of 6 Mc/s, I_{ant} is less than 1 dB below the maximum value.

Modulation curves

In fig. 19 the curves of various currents plotted against the alternating input voltage V_i are reproduced, viz. the "aerial current" I_{ant} , the anode (direct) current I_a , and the grid current I_g . Where $I_{ant} = 100\%$, $P_0 = 100$ kW. The curve for I_{ant} is the modulation characteristic for modulation applied to one of the preceding stages (modulated input voltage); the analogous curves for modulation of the output stage (V_i constant, V_g variable) are shown in fig. 20.

To ensure satisfactory contrast (γ) of the image as received, the curve for I_{ant} between black and white levels (75% and 10%) should be reasonably linear in both cases. Although nothing has yet been officially laid down as regards this linearity, it may be assumed that the curves as shown would be sufficiently linear. (It is to be expected that ultimate requirements will not be so stringent as

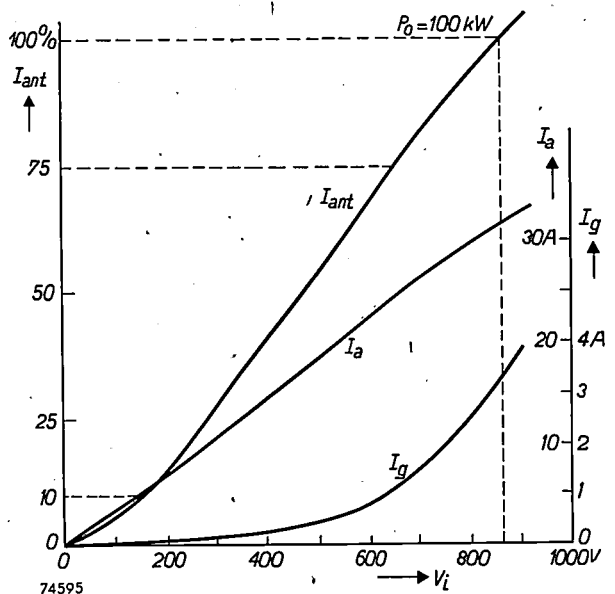


Fig. 19. Output curves for modulation in a preceding stage. "Aerial current" I_{ant} , anode (direct) current I_a and grid (direct) current I_g (I_a and I_g for the two TBW 12/100 valves together) plotted against the R.F. input voltage V_i (maximum peak between cathode and grid). D.C. grid voltage $V_g = -250$ V. Supply voltage $V_B = 6500$ V. Loading resistance $R_a = 350 \Omega$. Frequency $f = 48.25$ Mc/s (T.V. channel 2). $I_{ant} = 100\%$ corresponds to synchronisation peaks $P_0 = 100$ kW, $I_{ant} = 75\%$ (Cf. fig. 18) to black level and $I_{ant} = 10\%$ to white level.

those covering sound transmission, since the eye is very much less sensitive to non-linearity than the ear.) If necessary, the modulator can be designed with non-linear characteristics such that the non-linearity in the modulation curve is largely compensated.

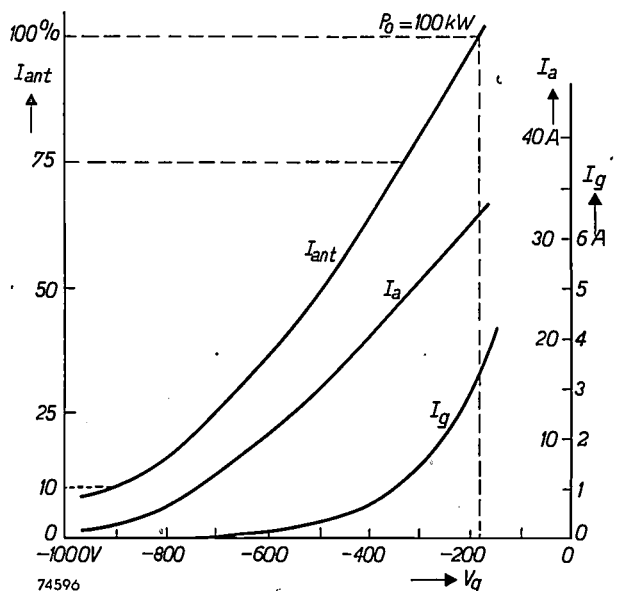


Fig. 20. Output curves for modulation in the output stage. I_{ant} , I_a and I_g as function of V_g , with $V_i = 770$ V, $V_B = 6500$ V, $R_a = 350 \Omega$, $f = 48.25$ Mc/s. $I_{ant} = 100\%$ again corresponds to $P_0 = 100$ kW.

From the anode current curves in figs 19 and 20 it is seen that at $P_0 = 100$ kW the anode takes a current of 32 A, which, at 6500 V D.C., represents 210 kW. The grid current curves show that the control stage has to deliver current peaks of considerable amperage.

Summary. An experiment has been carried out to see whether a television transmitter can be designed for high power and having adequate bandwidth for a system of 625 lines. The object was to obtain, by means of an output stage consisting of a single circuit, a resonance characteristic which would be flat within 1 dB throughout a range of 6 Mc/s, using equipment intended to deliver 100 kW peak.

A description is given of the experimental equipment, in

which the frequency is variable between 48 and 68 Mc/s (approx.); this range covers three television channels of 7 Mc/s band-width each. The equipment is capable of delivering 100 kW continuously, this being regarded as the peak value of the power that would occur during transmission of the synchronisation signals, with negative modulation. Resonance curves were plotted in respect of the corresponding black level, and these were found to meet the requirements imposed. Modulation curves were also plotted, for modulation in the output stage and in preceding stages, and these curves are reasonably linear between black and white levels.

The installation consists of an output stage for 100 kW and includes two TBW 12/100 valves in a push-pull earthed-grid circuit, preceded by a 20 kW driver stage with six type TBW 6/6000 valves, 2 sub-driver stages, and a low-power oscillator with variable frequency. The output stage was loaded with a water-cooled resistor, and the value of the power dissipated in this resistor was derived from the increase in temperature of the cooling-water.

A WEST-EUROPEAN TELEVISION NETWORK ON THE OCCASION OF THE CORONATION CELEBRATIONS IN LONDON

621.397.743

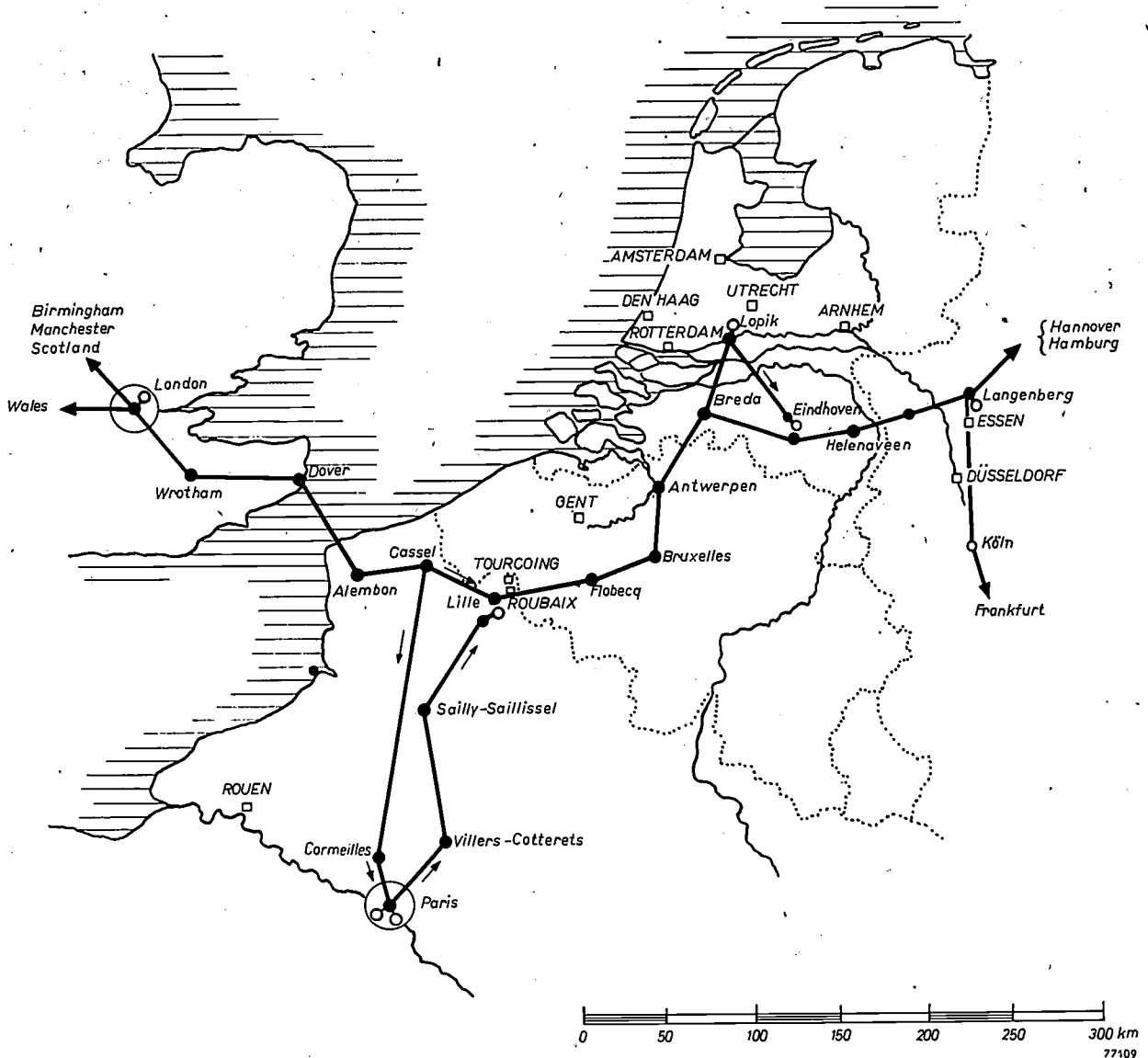


Fig. 1. Map of the provisional European television network. The circles represent television transmitters, the black dots radio-link transmitters and receivers.

The Coronation of Queen Elizabeth II on 2nd June 1953 was the occasion for the most ambitious television broadcast yet attempted in Europe. For this event a provisional international network was set up over which the British programme could be relayed to France, the Netherlands and West Germany.

The network itself (*fig. 1*) was made up of a chain of radio-links which relayed the vision signals to the various television transmitters. The sound signal was sent by land-line.

Originating in London, the signal was relayed

in four steps to Cassel and thence in two further stages to Paris. Some of this part of the network had been already set up for previous transmissions between London and Paris. From Cassel a branch chain of radio-links took the signal to Breda in Holland, the final link with the Dutch television transmitters at Lopik and Eindhoven. From Eindhoven the signal was sent via a further link station (Helenaveen) to Germany. The radio-link equipment from Lille to Lopik and Helenaveen was provided by Philips Telecommunication Industries of Hilversum.



Fig. 2. Trailer housing the 405-625 "line-converter", at the foot of the church-tower in Breda on which a radio-link receiver and two transmitters are installed.

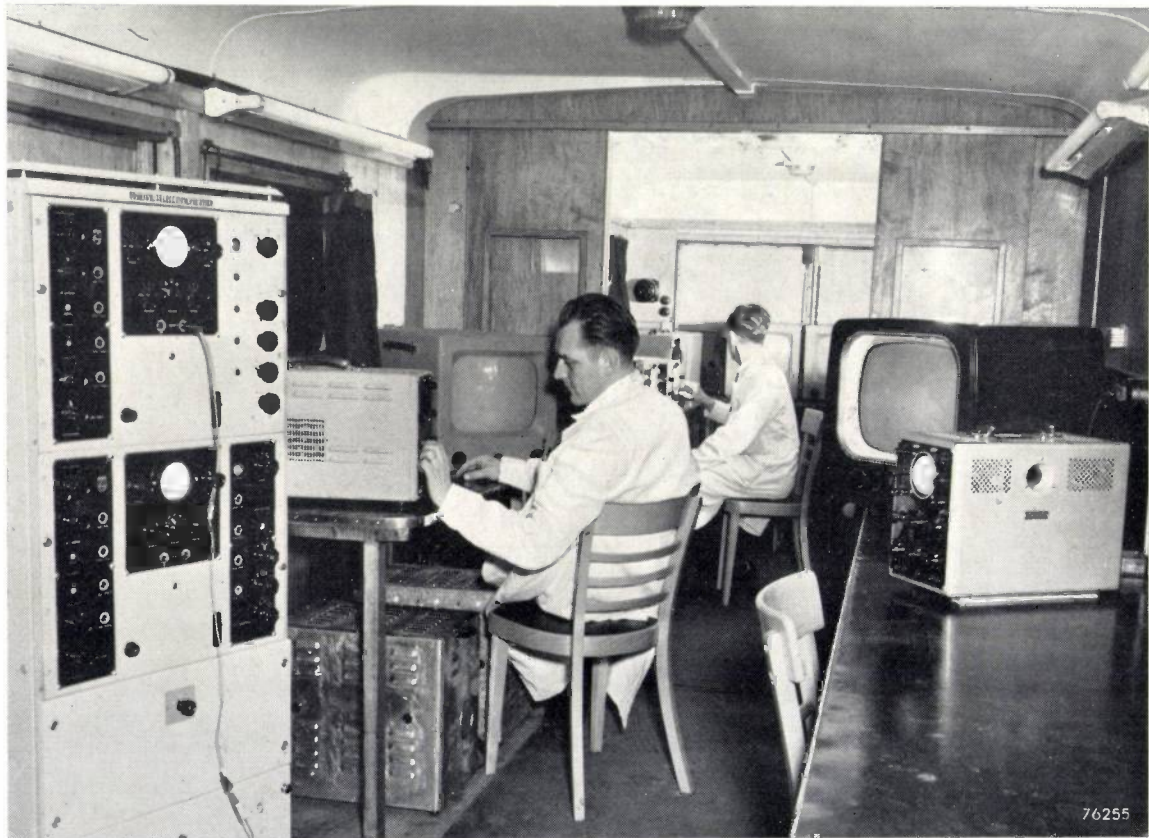


Fig. 3. Interior of the trailer stationed at Breda. The line-converter, not clearly visible in this photograph, is at the rear. Next to it are monitors for the 405 and the 625-line pictures. At the left in the foreground is the cabinet containing the generators of the line and frame frequencies for the 625-line system. The frame frequency was derived from the line frequency by a frequency divider, and synchronised with the frame frequency of the British system.

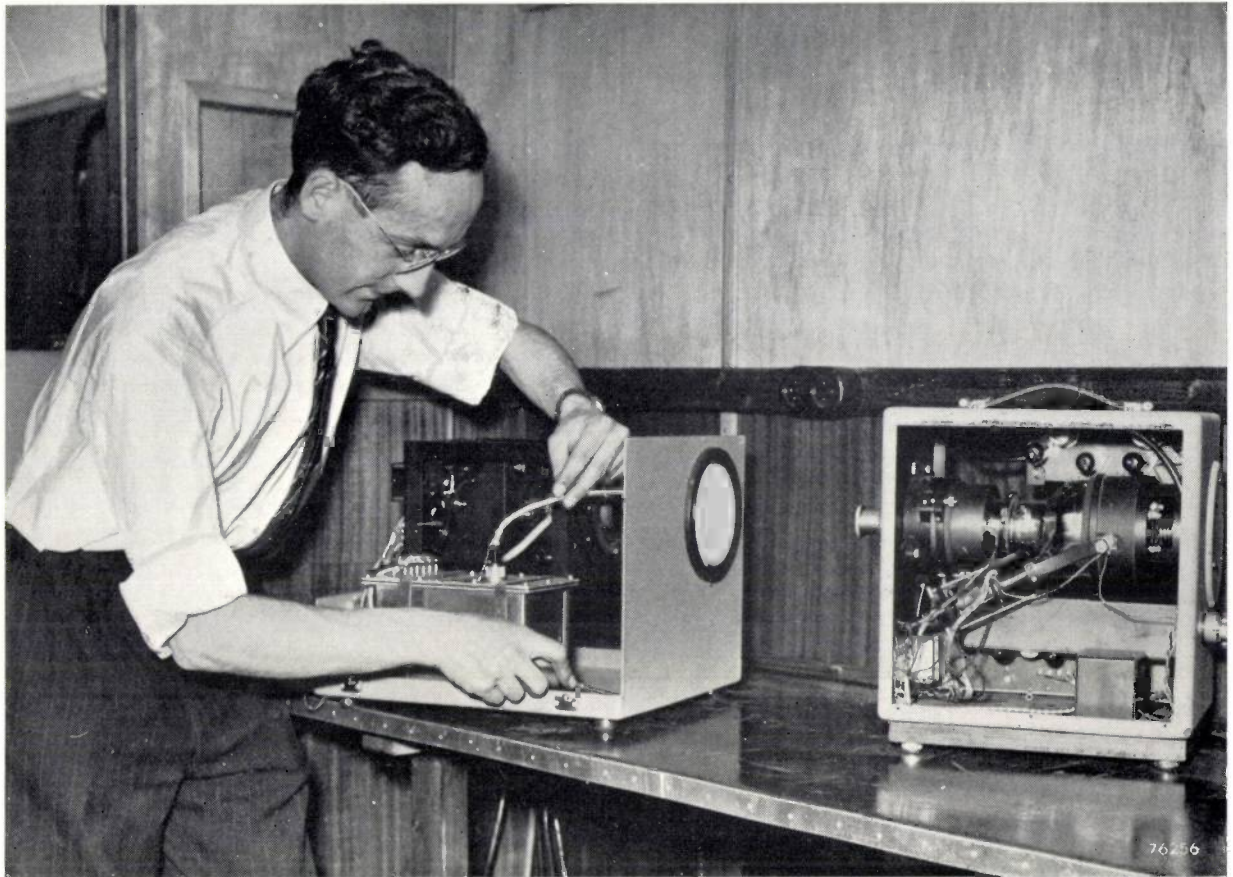


Fig. 4. The 405-625 line-converter used at Breda. On the left is the unit containing the picture-tube on which the 405-line picture is displayed, and on the right is the camera with the image-iconoscope in which this picture is scanned in the 625-line system.

Two points in the network are particular interest. These are the stations at which the British system (405 lines) is converted into the French (819 line) and the Dutch and German (625 line) systems. The conversion into these two systems was effected by "line-converters" located at Paris and at Breda.

The line-converter at Breda, designed at the Philips laboratories in Eindhoven, was installed

in a specially equipped trailer (*fig. 2*); the interior of this trailer is shown in *fig. 3*. The essentials of the line-converter are shown in *fig. 4*. The 405-line picture is displayed on a picture-tube and taken up by a special television camera which scans it in the 625-line system. The principle is thus very simple but precautions are necessary to suppress various disturbing effects. Further details of this apparatus will shortly be published in this Review.

A SHORT LENGTH DIRECT-VIEW PICTURE-TUBE

by J. L. H. JONKER.

621.385.832:621.397.62

The design of cathode-ray tubes for television reception is in a continuous state of development. Progress is usually directed along conventional lines, but occasionally an unorthodox approach yields fruitful results. This article describes a novel approach to the problem of shortening the length of picture tubes; although this development has not been accepted for production purposes, it appears to work surprisingly well.

The original cathode-ray tubes employed in television receivers were no different from those which had for some time previously been used for oscilloscopes. It was not long, however, before these were no longer able to meet the increasingly high standards set for television reception, for which a separate class of cathode-ray tubes came to be developed, known as television picture-tubes (in America: "kinescopes").

As far as these requirements are concerned, the public demand has been (and still is) for the largest possible picture compatible with a receiver of reasonable dimensions and price. As to the dimensions, designers of television sets request that the tube shall be as small (particularly as short) as possible; furthermore, to reduce costs, they prefer to use the lowest possible amount of power for focusing and deflection of the beam. Also, it must be possible to mass-produce the tube in such a way as to ensure uniform characteristics.

In recent years there has been no lack of evidence of the general trend in the development of picture-tubes. Although the tubes remain basically cathode-ray tubes, the dimensions and form have undergone such modification that the requirements referred to above can now be more closely satisfied. Under pressure of public demand, larger and larger screens have been made^{1,2}; in the United States, for example, screens have now reached a size of 75 cm diagonal. If it were not for the fact that the ratio of tube length to screen width has also undergone a radical decrease in the meantime, it would have been necessary — even with medium-sized pictures — to make TV cabinets so deep that they would have represented a real obstacle in the average

living-room. Thus reduction in the length of picture-tubes has been one of the main objects of the designer in the last few years.

Means of shortening the tube

The length of a tube can be divided roughly into three parts (fig. 1), viz. the lengths l_1 of the neck and l_2 of the cone, and the depth l_3 of the screen, which has to be curved to withstand the pressure of the atmosphere.

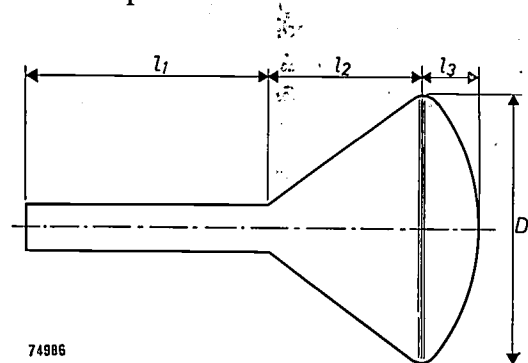


Fig. 1. The length of a television picture-tube is the sum of the neck length (l_1), the cone length (l_2) and the depth of the screen (l_3). D is the diameter or diagonal of round or rectangular screens respectively.

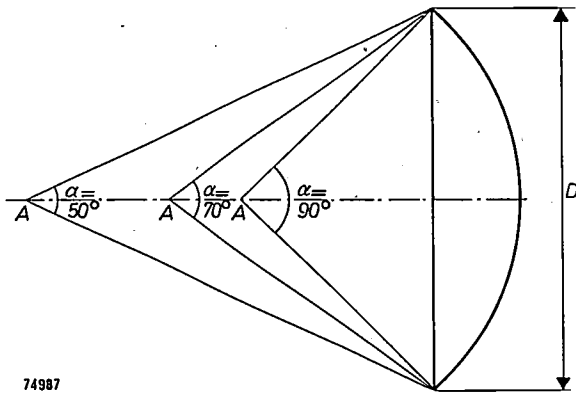
Every effort has been made to reduce the length of each of these parts in relation to the diagonal D ; by achieving the maximum compactness of the electron gun and the focusing and deflection coils, it has become possible to effect some reduction in the length of the neck; we shall refer to this again presently. The depth of the screen (l_3) has been reduced in those tubes of which the cone is made of metal instead of glass, as a smaller curvature is then practicable³). The most important reduction, however, concerns the cone. Any reduction in the length of the cone will of course be accompanied by an increase in the angle of deflection α , which is

¹) Round screens are steadily giving way to rectangular screens, thus saving the space occupied by those segments of the round screens which the rectangular picture does not utilize.

²) Another method of securing a large image is by projection (see Philips tech. Rev. 10, 69-78, 1948), but this is not within the scope of the present article.

³) J. de Gier, Th. Hagenberg, H. J. Meerkamp van Embden, J. A. M. Smelt and O. L. van Steenis, A steel picture-tube for television reception, Philips tech. Rev. 14, 281-291, 1953 (No. 10).

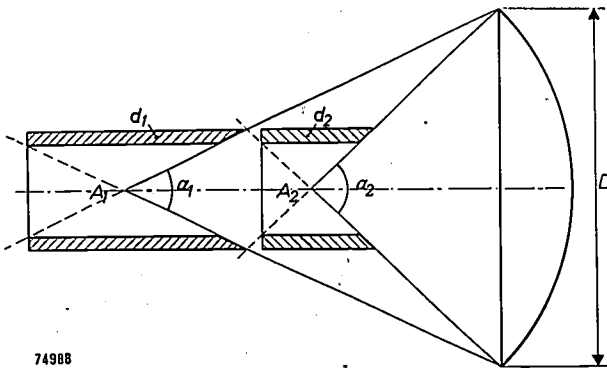
the angle between the extreme limits of the deflected beam (*fig. 2*). Step by step this angle has been increased⁴⁾ from 50° to 70° or even 90°, and this has necessitated a greater number of ampere-turns for the deflection; more, in fact, than would be proportional to the increase in the angle α . This results



74987

Fig. 2. If the cone be shortened, D being constant, this involves an increase in the deflection angle α . In older tubes this angle was about 50°, but in modern tubes α is 70° to 90°.

from the fact that the point about which the beam pivots when deflected lies in the centre of the deflection coils; if the deflection angle α is increased, the deflection coils must be shorter, as will be seen from *fig. 3*. This means that the electrons are subjected to the deflecting field over a shorter distance, i.e. that an increase in the strength of this field in proportion to α is not sufficient.



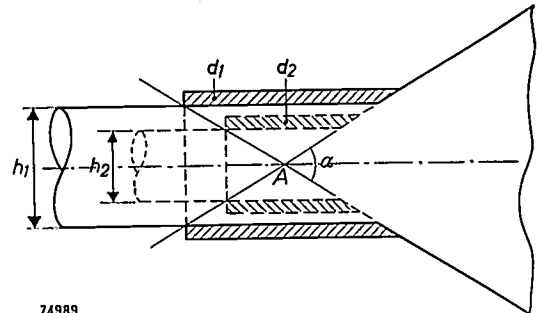
74988

Fig. 3. If the deflection angle be increased from α_1 to α_2 , the deflection coils (d_1 and d_2) must be shorter, since the pivoting-point of the beam (A_1 , A_2) must always be in the centre of the coils.

In general, a large number of ampere-turns will demand a greater amount of power. However, means have been found to reduce the amount of power required, the more important of these being as follows.

⁴⁾ L. E. Swedlund and H. P. Steiner, Short 16-in. metal-cone kinescope development, *Tele-tech* 9, 40-43 and 59-60, Aug. 1950. H. W. Grossböhl, The design of 90° deflection picture tubes, *Tele-tech* 10, 42-44, Aug. 1951.

- 1) So-called economy circuits have been devised for producing the deflection current, whereby a large part of the magnetic energy that accumulates in the deflection field can be recovered and fed back into the supply⁵⁾.
- 2) Special amplifying tubes (e.g. PL 81 and PL 82) ensuring higher efficiency, have been designed for these economy circuits.
- 3) Losses in the deflection coils have been reduced, e.g. by the use of Ferroxcube.
- 4) Economy in the power required for deflection can be achieved in the first instance by reducing the neck diameter of the tube, as far as is compatible with mechanical strength. As will be seen from *fig. 4*, however, a reduction in the thickness of the neck leads to shorter deflection coils. As pointed out, a short coil requires more ampere-turns, and the saving is therefore less than anticipated; but this effect can be largely counteracted by making the cone and neck merge into each other gradually in a certain manner and adapting the deflection coils to the resultant contour⁶⁾.



74989

Fig. 4. A reduction in the neck diameter from h_1 to h_2 means that the deflection coils (d_1 , d_2), for the same angle α , must be shorter.

However, all these devices do not alter the fact that the amount of power needed for deflection purposes constitutes a limitation on the size of the angle α . And this is not the only limitation. As the angle α is made larger, two further adverse effects become apparent.

As the cone is shortened, i.e. as the pivoting-point of the beam (A , *fig. 2*) is displaced in the direction of the screen, the surface defined by the focus on deflection will differ more from the surface of the screen, this being accompanied by greater fluctua-

⁵⁾ See J. Haantjes and F. Kerkhof, *Philips tech. Rev.* 10, 307-317, 1949. An article on new economy circuits will later be published in this Review.

⁶⁾ C. V. Bocciarelli, Low-power deflection for wide-angle C-R tubes, *Electronics* 25, 109-111, Sept. 1952.

tion in the size of the light spot. In order to limit this defocusing effect as much as possible, the greatest attainable depth of focus of the beam must be aimed at, i.e. the beam should be as narrow as possible ⁷⁾).

The other effect experienced with increasing α is related to "pin cushion" distortion. This occurs when the surface defined by the focus does not coincide with the surface of the screen. It is corrected by making the deflection field non-homogeneous — the field being weaker at points away from the axis. However, when the beam passes through an inhomogeneous field, a certain amount of aberration is introduced (which increases with the deflection) as a result of the unequal deflection of different parts of the beam. The narrower the beam, therefore, the less the aberration.

Both these effects can be counteracted only by using a narrower beam. Apart from refinements in the focusing, this can be done only by reducing the beam current (and hence the brightness of the picture), or by increasing the anode potential, which entails higher costs. A practical limit on beam attenuation, and thus also on the angle α , is accordingly soon reached.

The conclusion, therefore, is that any increase in the deflection angle may be accompanied by a useful reduction in the length of the tube, but also involves less desirable effects which in turn have to be remedied. It seems doubtful whether the angle α will be made much larger than 90° , which is already quite large.

Let us now consider the neck of the tube once more. It has been found possible to shorten this to some extent by reducing the length of the deflection coils and, again, by using a shorter electron gun. As regards the latter, two of the components have undergone some modification in the last few years, viz. the focusing system and the ion trap; we shall now take these in turn.

The focusing system

The electron beam can be focused by means of a magnetic lens, an electrostatic lens, or a combination of both. In picture-tubes the first-mentioned method is the more usual.

The reduction in the cone length referred to above has brought the focus nearer to the lens; in other words, the image distance has been reduced and the lens had to be made more powerful. In any shortening of the neck, i.e. reduction in the object distance, the lens must also be stronger. If the magnetic

field is generated by a coil, a stronger lens means more energizing power and this is not readily acceptable.

Different ways out of the difficulty have therefore been sought, and one of these consists in combining the magnetic lens with electrostatic pre-focusing: a less powerful lens is then necessary. This has the disadvantage, however, that the stray field of the magnetic lens, at the point where the electrostatic lens is situated, must be negligibly weak; otherwise, if the centering is not perfect, astigmatism occurs. This necessitates a certain neck length, which partially nullifies the reduction in the length already obtained.

A similar difficulty is encountered in wholly electrostatic focusing, to which much attention has been given in recent years with a view to economy in materials ⁸⁾. Here it is the stray field of the deflection coils that has to be kept away from the electrostatic lens. A special form of construction is therefore necessary which, in turn, once more increases the length of the neck.

Perhaps the best method is one that makes use of permanent magnets; no energizing is then necessary, and a considerable quantity of copper is saved. Furthermore expensive magnet steels can now be replaced by the non-metallic material Ferroxdure, which contains no scarce materials such as cobalt or nickel ⁹⁾.

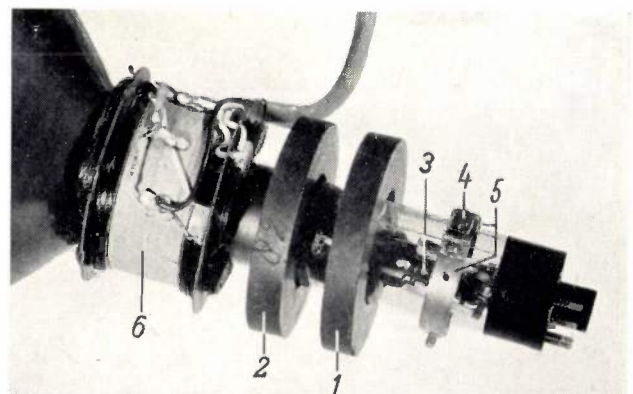


Fig. 5. Picture-tube in which the electron beam is focused by two ring-shaped magnets (1, 2) of Ferroxdure. These rings are magnetized in the axial direction of the tube and are mounted with like poles facing each other. 3 electron gun with ion trap comprising a small steel magnet (4) and two pole pieces (5). 6 deflection coils.

⁷⁾ The photograph on p. 368 of this issue illustrates a method of checking the deflection defocusing.

⁸⁾ L. E. Swedlund and R. Saunders, Material-saving picture tube, *Electronics* 24, 118-120, April 1951.
 C. S. Szegho, Cathode-ray picture tube with low focusing voltage, *Proc. Inst. Rad. Engrs.*, 40, 937-939, 1952 (No. 8).
 C. T. Allison and F. G. Blackler, A univoltage electrostatic lens for television cathode-ray tubes, *Conv. Brit. Contrib. Telev.*, 1952, art. R8-1333.
⁹⁾ J. J. Went, G. W. Rathenau, E. W. Gorter and G. W. van Oosterhout, *Philips tech. Rev.* 13, 194-208, 1952 (No. 7).

Fig. 5 illustrates the method of focusing by means of Ferroxdure magnets; two flat rings of Ferroxdure are used, these being magnetized in the axial direction and fitted to the neck of the tube with like poles facing each other. The desired configuration of the field is obtained by varying the space between the magnets. The high coercive force of Ferroxdure renders this material very suitable for magnetizing in the direction of the thickness when made in the form of flat rings¹⁰⁾ and, as we wish to keep this thickness as small as possible, the ring-shaped magnet serves the purpose well.

The ion trap

When the picture-tube is operating, negative ions are produced which originate from residual gases, or from the cathode. Now, since the mass of an ion is very much greater than that of an electron, the field of the deflection coils has very little effect on the ions and, if nothing is done to check them, they all strike the centre of the screen, where chemical action takes place and reduces the luminescence. After a time, a dark spot becomes visible in the centre of the picture.

To prevent this, an "ion trap" is incorporated in the electron gun, and this functions by reason of the circumstance just mentioned, that a magnetic field deflects ions much less than electrons; in an electric field, however, the degree of deflection is the same. Fig. 6*a* illustrates the principle of the ion trap. An electrostatic lens is employed to deflect the paths of the electrons and ions from the axis. Further along, successive deflections by two permanent magnets bring the electron beam back to the axis, whilst the ions strike the second acceleration electrode (the anode), where they are rendered harmless.

A reduction in the length of the neck can be achieved by mounting the first part of the electron gun — viz. the cathode, the control grid and the first acceleration electrode — at an angle with the main axis (fig. 6*b*, fig. 5); this dispenses with the need for one of the deflections, and one of the magnets thus becomes superfluous.

The induction which the ion-trap magnet has to provide is 3.5×10^{-3} to 6×10^{-3} Wb/m² (35 to 60 gauss), and a small steel magnet is generally used (4, fig. 5).

One function of the ion trap, not implied in the name but just as important as the neutralizing of ions, is the means of accurately centering the beam when once the deflection coils and focusing system

have been mounted on the tube; in the absence of such centring the oblique passage of the beam through the deflection coils produces distortion of the image. The ion trap magnet thus permits correction of residual errors in the centering.

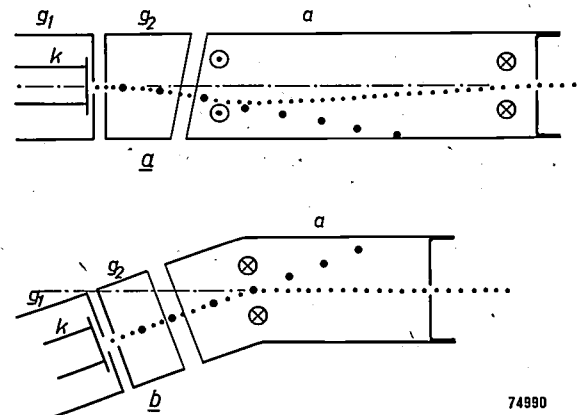


Fig. 6. Old (a) and new (b) form of ion trap. *k* cathode; *g*₁ control grid (Wehnelt cylinder); *g*₂ first acceleration electrode (low positive potential); *a* second acceleration electrode (anode, at high positive potential). Small dots: electrons. Large dots: negative ions.

Action in (a): The oblique gap between *g*₂ and the anode *a* imparts to the field a component perpendicular to the axis of the tube, which deflects the paths of electrons and ions an equal extent from the axis. A magnetic field ⊙ perpendicular to the axis returns the electrons to the axis, but has practically no effect on the direction of the ions, which strike the anode and are thus rendered harmless. A second magnetic field ⊗ opposed to the first, deflects the electrons, so that they move parallel to the axis.

Action in (b): *k*, *g*₁, *g*₂ and the first section of the anode *a* are here mounted with their axis at a certain angle to the axis of the tube. Only one magnetic field (⊗) is required to bring the path of the electrons parallel to the axis. The ions are hardly affected by this field and again fall on the anode.

Tube with bent neck

In the foregoing we have outlined some of the obstacles encountered when efforts are made to reduce the length of picture tubes in their present form. Some of these obstacles will doubtless be surmounted in due course, but we will not speculate on such possibilities here. Instead, we shall describe a tube which is considerably shorter than those of the conventional type and which was constructed as an experiment in the Philips laboratories at Eindhoven.

Arising from the effect of the ion trap, in which the beam is given a permanent deflection, the idea was conceived of bending the neck of the tube through an angle of 90° or more and making the beam follow the curve by means of a magnet. In this way the length of the tube can be appreciably reduced (fig. 7), the bend serving simultaneously as an ion trap¹¹⁾.

¹¹⁾ Proc. Inst. Rad. Engrs. 36, 1485, 1948 (fig. 4) contains an illustration of a tube whose neck is bent slightly, solely to produce the effect of an ion trap.

¹⁰⁾ See article referred to in footnote 9), p. 196.

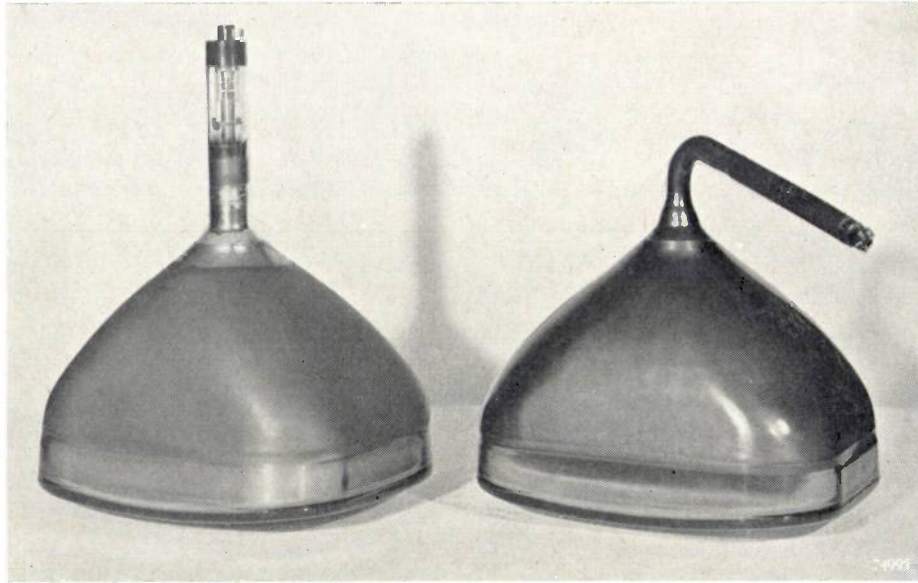


Fig. 7. Left: Conventional picture-tube (type MW 36-22). Right: Experimental tube with bent neck to reduce the over-all length of the tube.

The most suitable point for the bend is between the focusing system and the deflection coils, the latter being then so constructed that they can pass over the bend; this has, in fact, proved quite practicable.

For curving the beam a small permanent magnet was employed (figs 8 and 9) having an induction of roughly 7×10^{-3} Wb/m² (= 70 gauss). Adjustment of the position and strength of the field (the

strength by means of a magnetic shunt) centres the beam exactly on the axis of the deflection coils in the same way as does the ion trap. To avoid astigmatism, the deflection field must be symmetrical about the plane bisecting the deflection angle, and this is not difficult to achieve. A narrow beam is again an advantage.

In the construction of the experimental tube depicted in fig. 7 we used as many as possible of

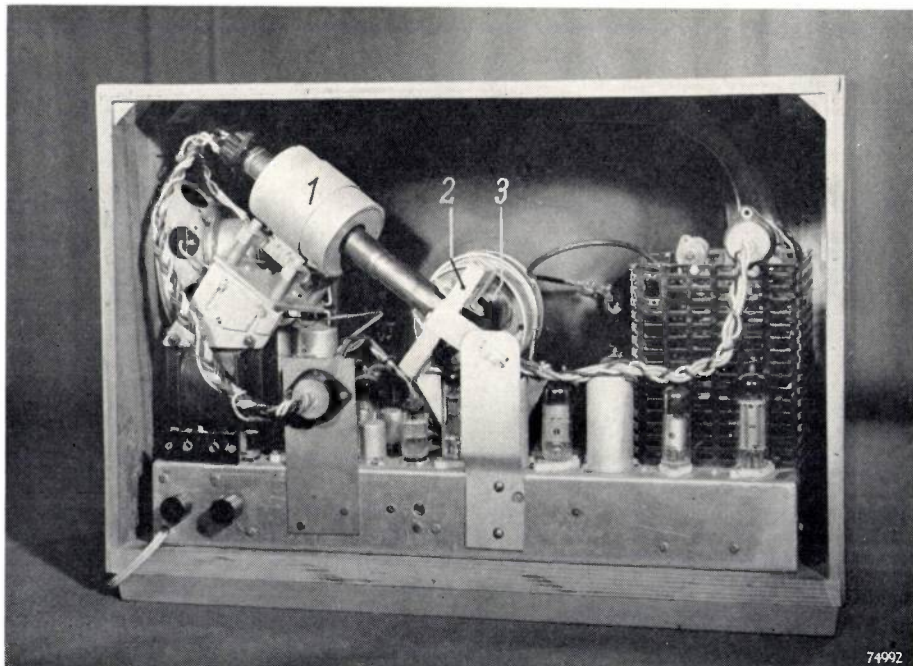


Fig. 8. Tube with bent neck mounted in a television receiver. 1 focusing coil; 2 deflection magnet to make the beam follow the curve in the neck; 3 deflection coils.

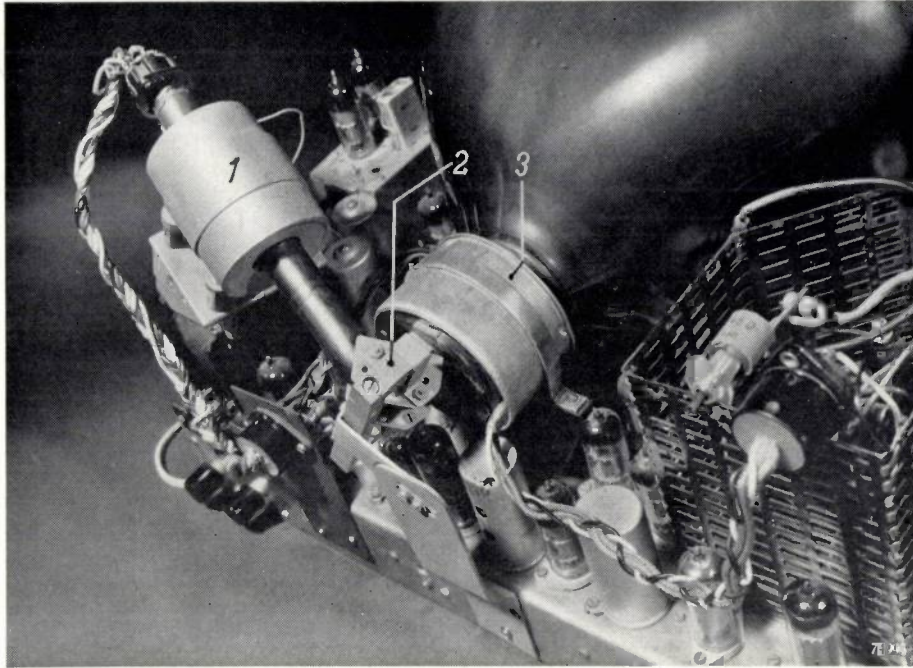


Fig. 9. Close-up of the tube neck and deflection coils.

the standard components of tube MW 36-22, viz. the screen (25 cm \times 32.5 cm), the cone, the electron gun (without ion trap) and the deflection coils. The deflection angle (α) is the one usually employed, viz. 65° ; the normal deflection power is accordingly sufficient. Only the focusing coils and the permanent magnet were specially made.

As the length of the bent part of the neck in no way affects the depth of the television cabinet, it can, if desired, be longer than that of a tube with straight neck; the neck of the experimental tube is, in fact, longer as this ensures better focusing: the distance from object to lens is greater, the magnification correspondingly less and the light spot smaller. At the same time, a long focusing coil can be used, the diameter of which will then be smaller. By bending the neck through an angle larger than 90° we ensure that the focusing coil lies within the over-all length of the tube, so that only this length determines the depth of the receiver cabinet, not the tube length plus the thickness of the coil.

The direction in which the neck is bent is immaterial; in this tube the direction lies in the plane of the diagonal of the screen, for reasons of available space in the cabinet (fig. 8).

A neck bent in the horizontal plane also has certain advantages. Such an arrangement would facilitate supporting the focusing coil and the permanent magnet on the chassis. A second advantage is related to the fact that, when a permanent magnet is used for bending the beam, the picture is displaced slightly on the screen when the high tension varies. With the neck of the tube bent horizontally this displacement would also

be horizontal, and this would not be so noticeable as diagonal displacement, as is now the case. It may be added that such displacement can be entirely avoided by employing a stabilized high tension supply, the output of which is constant in spite of variations in the mains voltage or in the load¹²).

Using this tube a television receiver was constructed (fig. 10), the over-all dimensions of which were as follows:

width	20"
height	14 $\frac{1}{2}$ "
depth	13 $\frac{3}{4}$ "

i.e. no larger than a medium-sized radio receiver.

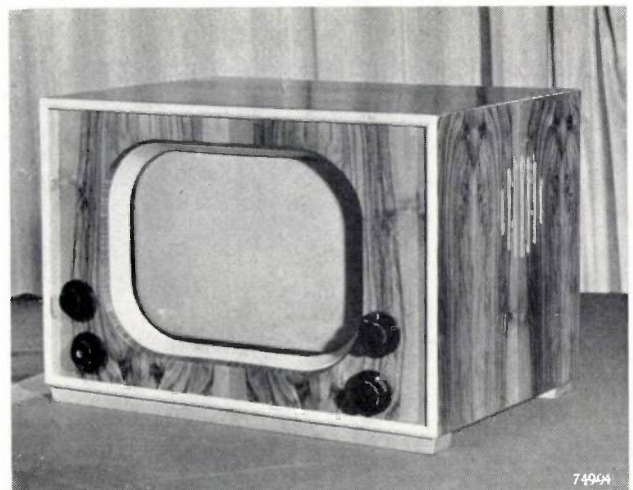


Fig. 10. Front view of the receiver illustrated in figs 8 and 9. The loudspeaker is at the side.

¹² See for example, J. J. P. Valetton, Philips tech. Rev. 14, 21-32, 1952 (No. 1).

The height has been kept down by mounting the loudspeaker at the side of the cabinet instead of below the tube, a method that is quite often employed (see also fig. 8).

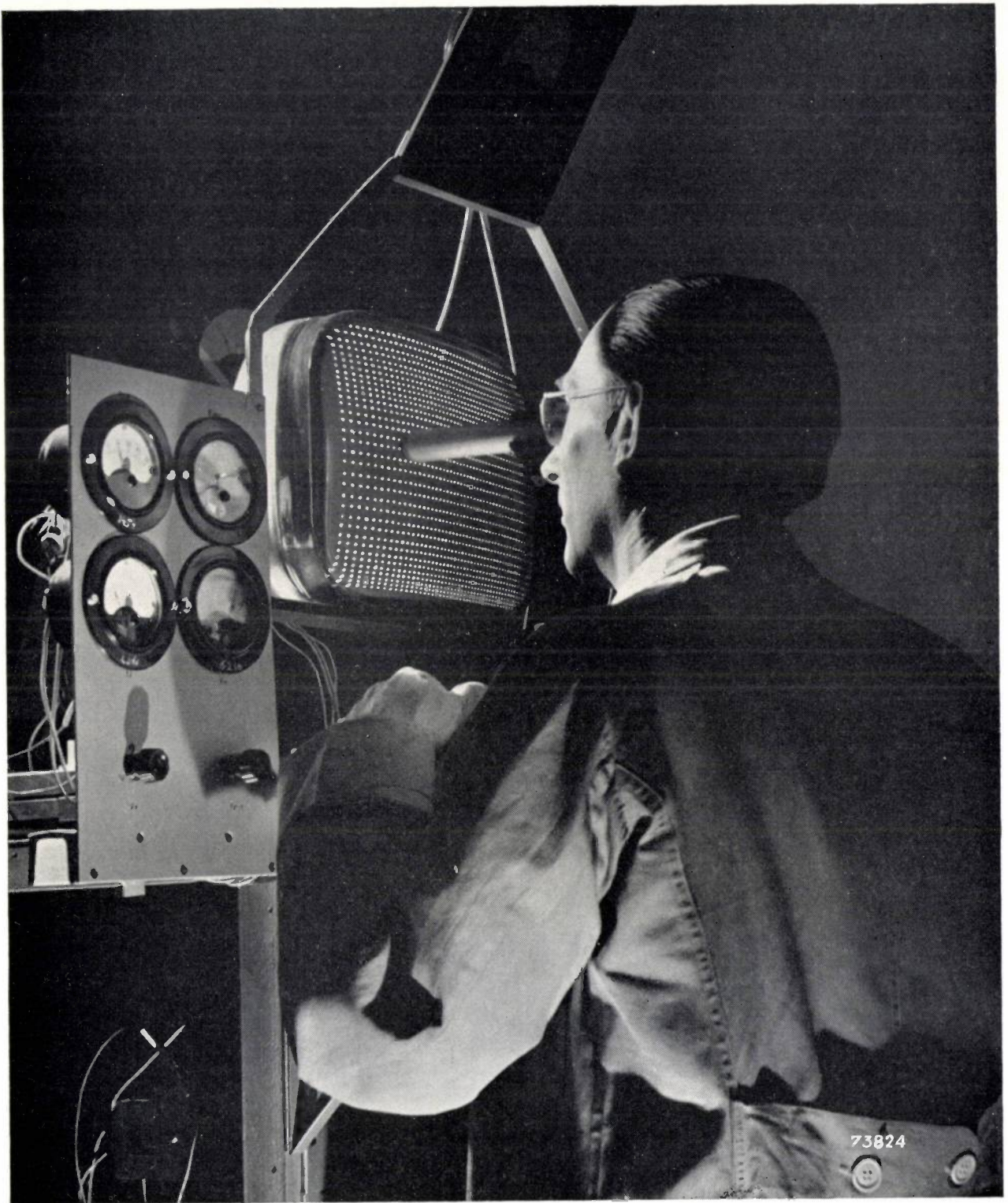
It cannot be pretended that the bent-neck tube is more than an experimental idea, and it should not be expected that such tubes will be put into production for the present. Though practical application may be delayed, however, it was considered that publication of a description of this laboratory model was not inopportune.

Acknowledgements for their co-operation in the design of both tube and receiver are due to Messrs. Dammers, Diemer, Neeteson and De Weyer.

Summary. Arising from the demand for larger television pictures, the development of direct-vision tubes has been directed towards larger and larger screens. This has of course resulted in greater tube lengths and, as this length determines the depth of the receiver cabinet, television sets have shown a tendency to become rather unwieldy. In recent years, therefore, every effort has been made to reduce the length of the tube in relation to the picture dimensions. Some of the methods employed are outlined, viz. shortening the cone (accompanied by wider deflection angle), reducing screen curvature (as in the metal-coned tube), focusing by Ferroxdure permanent magnets, and the use of simplified ion traps.

A description follows of an experimental tube whose neck is bent through 90° or more, resulting in an appreciable reduction in the over-all tube length. A permanent magnet is used to make the beam follow the curve in the neck. A television set was built for this tube, the depth of which was only $13\frac{1}{2}$ " (less than the diagonal of the screen); the width and height are 20 " and $14\frac{1}{2}$ " respectively. The deflection angle is the conventional 65 degrees.

CHECKING THE LUMINOUS SPOT IN CATHODE-RAY TUBES



Photograph Walter Nürnberg

The quality of a television picture is largely dependent on the sharpness of the luminous spot which builds up the image, line by line, on the screen of the cathode-ray tube — it depends, thus, on the accuracy with which the electron beam is focused to a point on the screen. The above photograph shows the inspection of spot size and shape during the manufacture of picture tubes. The tube is set up under normal operating

conditions and fed via the grid with a series of voltage pulses in such a way that a raster of spots is produced on the screen. The voltage pulse magnitude is so chosen that the amplitude of the pulsed electron beam current obtains a particular value e.g. $100 \mu\text{A}$. After careful adjustment of the focusing current, a microscope is used to determine whether or not the spots are small enough and sufficiently circular.

A BATTERY-OPERATED GEIGER-MÜLLER COUNTER

by G. HEPP *).

621.317.7:539.16.08:621.353:621.387.424

The development of counter tubes during the last few years has been characterized by their adaptation to specific problems that have arisen in scientific and technical applications. Today, in consequence, counter tubes are employed for a variety of purposes in quite simple equipment. In many of these applications small and conveniently portable units are desirable, and the following is a description of a Geiger counter for battery operation which, by reason of its small dimensions and low weight, will prove a valuable instrument in a wide range of uses.

The widening field of applications of radio-active substances and equipment that emits ionizing radiations (X-ray tubes, cyclotrons, betatrons etc.) is creating a growing demand for simple, easily-handled instruments for detecting and measuring radio-active radiations. Geiger counters commonly form the basis of instruments for this purpose. Such instruments are used amongst other things for locating radio-active ores and for the protective monitoring of radiations in laboratories. An example of a more technical application is the measurement of liquid level by means of a float containing a radio-active substance, the height of the float being ascertained with the aid of a counter. (The advantage of this method over numerous others is that it is not necessary to make any connections through the wall of the vessel.)

In the following, a description is given of a Geiger counter suitable for battery operation. As the current consumption is very low, small batteries can be used, of the kind employed in hearing aids, and the instrument is accordingly quite compact. It is fitted with a Geiger-Müller counter tube in which the quenching gas is a halogen¹⁾.

Working principle

The measurement of radiations by means of counter tubes generally takes place along the following lines. Every discharge in the counter tube produces a voltage pulse across a resistor in series with the tube, and these pulses are fed to a circuit which converts them into current pulses whose amplitude and duration are independent of the intensity of the discharge in the tube. (The discharge intensity is dependent on the voltage applied to the tube, as also on the characteristics of

the tube itself. The first of these drawbacks could be overcome by stabilizing the supply voltage, but this would mean too great a drain on the batteries. Furthermore, in view of the second disadvantage, pulse-shaping is in any case generally preferable.)

The shaped current pulses are applied to a capacitor which is charged across a resistor, the discharging current being then measured with a D.C. measuring instrument. Seeing that the discharges in the counter tube occur at random, the capacitor charging current will be subject to fluctuations. However, if the time constant (RC) of the capacitor C and the discharging resistance R , is large compared to the average time between pulses, the charging current will be sufficiently smoothed to ensure a meter deflection independent of the fluctuating pulse frequency.

The supply voltage for the counter tube can be obtained in different ways, namely:

- 1) from a battery capable of producing the required high tension. This involves heavy and bulky equipment, with fairly high running costs for battery renewals.
- 2) From a vibrator operated on a low-tension accumulator. Such vibrators require a considerable amount of power to maintain the motion of the contact springs.
- 3) From an oscillator, fed from the same battery as that used for the pulse-shaping valve. At first sight, even this method would not appear to be very attractive, since it necessitates the use of an additional valve, with consequent increased load on anode and heater batteries. However, it is possible to eliminate both these objections by employing only one valve for the two functions concerned; the oscillator then operates only at those moments when a discharge occurs in the counter tube. This arrangement ensures very low anode current consumption and consequently a longer battery life.

*) Temporarily with Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil.

1) See N. Warmoltz, Geiger-Müller counters, Philips tech. Rev. 13, 282-292, 1952.

Economy circuits for a Geiger counter

The principle on which the Geiger counter works is described with reference to *fig. 1* which shows the simplified circuit diagram. The counter tube *GM*, in series with a resistor R_1 is connected in parallel

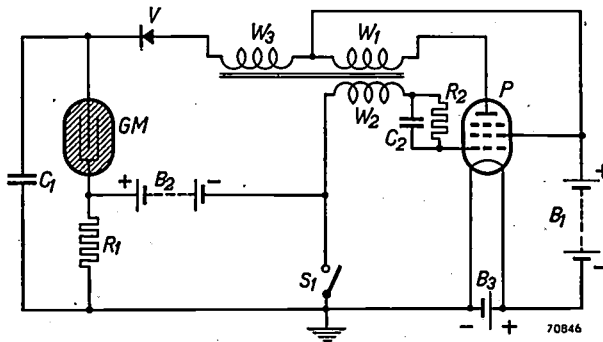


Fig. 1. Circuit diagram of battery-operated Geiger counter. *GM* Counter tube, *P* pentode, *V* rectifier element, B_1 H.T. battery, B_2 blocking battery, B_3 filament battery, S_1 starting switch.

with a capacitor C_1 which, under ordinary conditions, carries a sufficiently large charge to initiate an independent discharge of short duration in the counter tube on the arrival of an ionising particle. With each such discharge in the tube the capacitor C_1 loses some of its charge, and the voltage accordingly drops each time. After each discharge, the capacitor is re-charged to the requisite voltage via a rectifier element V ; the current pulse which is passed to the capacitor through the rectifier is supplied by a pentode P , connected as a blocking oscillator²⁾.

When switch S_1 is closed (so that the battery B_2 no longer affects the oscillator), current pulses occur at regular intervals in the anode circuit of the valve P . The time elapsing between pulses is governed mainly by the values of the capacitance C_2 and the resistor R_2 . The transformer, of which W_1 is the anode winding and W_2 the grid winding, includes a third winding W_3 , so connected that each current pulse in the anode circuit induces a pulse in W_3 which charges C_1 through V .

The function of the battery B_2 is to prevent the occurrence of current pulses in P when no discharges are taking place in the counter tube. This battery lowers the potential of the grid to a level with respect to the cathode such that oscillation cannot occur (with switch S_1 open). A discharge in the counter

tube produces a voltage pulse across the resistor R_1 which momentarily neutralises the voltage from the battery B_2 , and permits P to pass a current pulse.

As the valve P does not work continuously as a blocking oscillator, but only at those moments when the counter tube operates, the battery B_1 is not under continuous load and the circuit accordingly consumes very little current (the filament battery B_3 is of course loaded continuously, but the filament current of the valves used is only 13 mA at 1.5 V).

Since a discharge in the counter tube is essential for the occurrence of a charging pulse from the pentode and, as the discharge is possible only when the capacitor C_1 carries a certain charge, the circuit cannot start to function spontaneously. For this reason switch S_1 is provided to isolate the battery B_2 from the grid circuit and temporarily remove the block on the pentode P . When the instrument is first switched on, therefore, this switch must be closed for a moment, to enable the pulses to charge the capacitor C_1 to the required potential. When switch S_1 is opened again, each discharge taking place in the counter tube causes the charge on C_1 to be replenished in the manner already described, since the charge per pulse that can be applied to C_1 through V is greater than that utilized by the counter tube per count.

Unfortunately, however, the counter tube is not the only source of loss in charge of C_1 . For the lowest count rate, i.e. the "background" count (number of discharges per unit time in the absence of a specific ionizing event), the losses due to leakage in the capacitor itself, in the rectifier V and in the counter tube are much more significant. Consequently, unless C_1 receives an adequate charging current, the voltage across it will drop below the working voltage of the counter. As the instrument must continue to operate when once switched on, the leak currents must therefore be kept as low as possible. For this reason V cannot be a selenium or germanium rectifier, as the leakage of these is too high. A diode in the circuit shown in *fig. 1* would necessitate an additional filament battery, but a circuit which dispenses with the need for this is depicted in *fig. 2*. Here, instead of a positive voltage at the anode of the counter tube as in *fig. 1*, a negative voltage is produced at the cathode; in this circuit use is made of a diode D included in the envelope of the pentode. The capacitor which is charged by the current pulses is once more denoted by C_1 . The voltage pulses produced across the resistor R_1 are applied to the control grid of the pentode P through another capacitance C_2 , so that

²⁾ For the operation of a blocking oscillator see, for example, D. Goedhart and G. Hepp, Carrier supply in an installation for carrier telephony, Philips tech. Rev. 8, 137-146, 1946, and P. A. Neeteson, Flywheel synchronisation of saw-tooth generators in television receivers, Philips tech. Rev. 13, 312-322, 1952 (No. 11).

the pentode is cut off each time the counter tube has functioned. A special feature of the circuit illustrated in fig. 2 is that the potential occurring across the counter tube is the sum of the potential developed across C_1 and that of the battery B_1 ; it is thus not necessary for C_1 to develop the whole of the voltage required for the counter tube. If the voltage needed were, say, 360 V and that of the battery B_1 45 V, only 315 V would have to be developed across C_1 . A difficulty in connection with this circuit is that the leakage of capacitor C_1 is in part due to the finite insulation resistance of winding W_3 of the transformer; it has been found impossible in practice to attain a sufficiently high insulation to entirely eliminate this leakage.

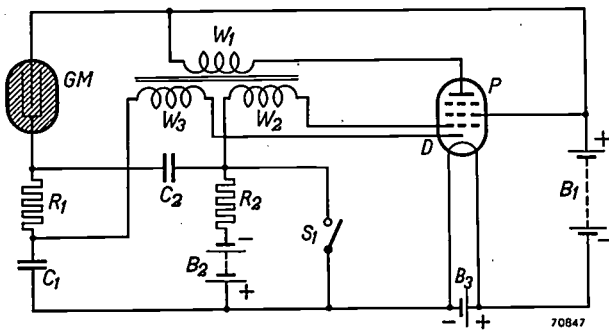


Fig. 2. Circuit diagram in which the rectifier shown in fig. 1 is replaced by a diode D included in the envelope of the pentode.

A simple calculation will determine the permissible order of size for the leak resistance. At a voltage of 360 V and with 1000 pulses per second, the mean value of the current flowing in the counter tube is about 1 μ A. A charge of 10^{-9} coulombs per pulse is therefore taken from the capacitor. If we denote the number of pulses per second by N , the counter tube will in that time consume a charge equal to:

$$N \times 10^{-9} \text{ coulomb .}$$

Further, if we denote the leak resistance of capacitor C_1 by R_1 , the leakage per second with 315 V across the capacitor will be:

$$\frac{315}{R_1} \text{ coulomb .}$$

With a very large number of counts per second taking place in the counter tube, the amount of the charge that the tube will take from the capacitor will be high compared with the loss due to the leakage current. For a low counting rate, however, the leak current gains in significance, since the charging pulses are proportional to the number of discharges taking place in the counter tube. Now, it is essential that, even at the lowest count rate, namely the background count, there will be sufficient surplus charge in the capacitor C_1 to compensate the leakage current.

Let I denote the maximum average anode current relating to a single charging pulse, T the duration of the pulse, and W_1/W_3 the ratio of primary to secondary (W_3) turns. The maximum charge that can be supplied to C_1 per second is then:

$$NI \frac{W_1}{W_3} T \text{ coulomb.}$$

To ensure that the potential across C_1 will not drop, this charge should be equal to or greater than

$$(N \times 10^{-9}) + \frac{315}{R_1} \text{ coulomb.}$$

If $I = 2$ mA, $T = 10^{-4}$ sec and $W_1/W_3 = 1/10$, $N \times 10^{-9} \ll NI (W_1/W_3) T$, so that:

$$NI \frac{W_1}{W_3} T \geq \frac{315}{R_1} .$$

Hence, if one discharge occurs every 5 sec ($N = 0.2$) in the "background", the leak resistance is required to be

$$R_1 \geq 10^{11} \Omega .$$

A circuit which imposes less strict requirements on the insulation resistance of the transformer is depicted in fig. 3, in which the capacitor C_1 is connected between winding W_3 of the transformer and the diode anode. In this kind of circuit great care must be taken to keep the capacitance between C_1 and the chassis as small as possible. As will be seen from fig. 3, this stray capacitance is in parallel with the self-capacitance of winding W_3 , these together being designated as C_p in the diagram. In order to raise the anode voltage of D to a level high enough to permit the charging of C_1 , the height of the pulse across W_3 must be at least equal to the potential across C_1 . C_p , which is charged repeatedly to this potential, must therefore be as small as possible. After each pulse, C_p discharges via the transformer winding. This charging energy is lost, at least in so far as the charging of C_1 is concerned. (It will be shown later, however, that this energy can be used for another purpose, i.e. to produce a deflection in the indicating instrument.)

The permissible magnitude of C_p can be estimated as follows: with the above-mentioned values of T , I and W_1/W_3 , the maximum available charge per pulse is 2×10^{-8} coulomb.

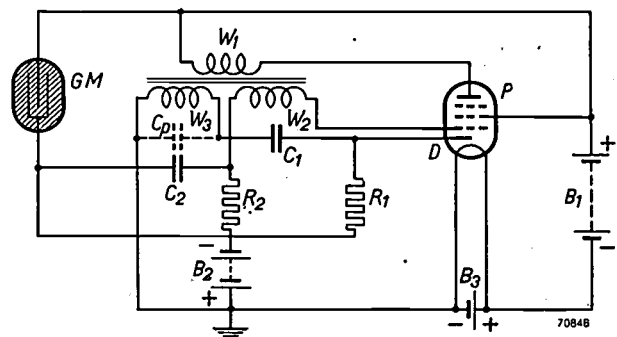


Fig. 3. Circuit diagram which imposes less stringent requirements on the insulation resistance of the transformer winding W_3 .

If C_1 and C_p are charged to a potential of 315 V, and not more than 10% of the charge is to be applied to C_p , the value of C_p must be such that:

$$315 C_p \cong \frac{2 \times 10^{-8}}{10}$$

or:

$$C_p \cong 6.3 \text{ pF.}$$

It has been found in practice that this can be accomplished by employing a special winding and insulating technique for the transformer, and by careful positioning of the components.

In order to keep the capacitance between C_1 and the chassis as low as possible, the physical dimensions of this capacitor must be limited, i.e. the value of the capacitor will also be low. A low capacitance is also desirable for another reason, viz that a short starting time is required, this being the time that switch S_1 must be kept closed for the capacitance C_1 to charge up (see figs 1 and 2). This period should preferably be limited to a fraction of a second and, of course, the lower the capacitance of C_1 , the shorter this will be. On the other hand, if the capacitor is too small, there is every risk that, with only a few pulses arriving within a given time, the voltage across C_1 will drop so much between pulses that the counter tube is unable to function, thus interrupting the working of the circuit. A sufficiently high value of C_1 is therefore used to minimise the chance that the circuit will cease to function, even with the smallest number of pulses (background) likely to occur in practice. The probability of failure due to this cause may be estimated numerically in the following way.

Suppose E_c to be the minimum voltage across the counter tube which makes the counting pulses large enough to remove the block on the oscillator. Taking E_n as the normal tube voltage, we introduce the quantity

$$p = \frac{E_c}{E_n}$$

If N_o is the number of pulses per second in the background and k the product of the capacity C_1 and the leak-resistance R_1 , then the time T_c necessary for the voltage E_n to decrease to the value E_c is derived from the relation:

$$E_c = E_n e^{-\frac{T_c}{R_1 C_1}} = E_n e^{-\frac{T_c}{k}}$$

from which
$$\frac{T_c}{k} = -\ln \frac{E_c}{E_n} = -\ln p$$

The chance that a given pulse is not followed within this time by a following pulse is

$$e^{-N_o T_c} = e^{k N_o \ln p} = p^{k N_o}$$

Of the $N_o T$ pulses occurring in a time T there are thus $N_o T p^{k N_o}$ which are not followed within the time T_c by another pulse. If the value of this expression is small in relation to

unity, it gives at the same time the probability that the critical voltage E_c will be reached within the time T .

It is tacitly assumed in this connection that the occurrence of a single pulse is sufficient under all circumstances to bring the voltage back to the value E_n . If, however, more than one pulse is necessary, the same conclusion is valid provided that the number of pulses required is relatively small.

If, for example, $k = 1000 \text{ sec}$, $p = 0.85$, and $N_o = 1/5$, then during a time T of one hour the chance that the voltage drops below the value E_c is:

$$\frac{1}{5} \times 3600 \times (0.85)^{200} = 4.5 \times 10^{-12}$$

Stumpers³⁾ has discussed this problem under the assumption that no limitation of the capacitor voltage is occasioned by the pulse amplitude, and that each pulse produces the same charge in the capacitor.

Finally, it should be mentioned that exceptionally high requirements are demanded for the self-inductance coefficient of the transformer windings. This must be very high in order to limit the magnetisation current. An estimate of the value required can be made with the aid of an equivalent transformer circuit.

This consists of an "ideal" transformer with windings of infinite self-inductance (leakage inductance and losses being thereby disregarded), and a coil, in parallel with one of the windings, whose self-inductance is equal to that of the winding of the actual transformer. Fig. 4 shows an equivalent circuit of

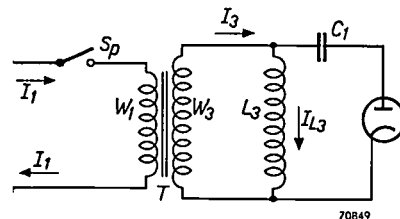


Fig. 4. Equivalent circuit diagram of the transformer. The pentode is represented by a source of current I_1 and a switch S_p .

this kind, based on a part of fig. 3. In this figure T represents the ideal transformer; the current I_3 is equal to $I_1 W_1 / W_3$. The pentode P is represented by a switch S_p , by means of which the current I_1 is periodically applied. Suppose that $I_1 = 2 \text{ mA}$ and that $W_1 / W_3 = 1/10$; then $I_3 = 0.2 \text{ mA}$. Immediately the circuit is closed, this whole current will flow in the capacitor C_1 as well as through the diode, but an increasingly large part of I_3 will also flow in the coil of which the self-inductance L_3 is the same as that of W_3 (fig. 3). If the voltage across C_1 is 315 V, the time derivative of the current in L_3 will be determined by:

$$L_3 \frac{dI_{L3}}{dt} = 315.$$

Suppose now that during the period of one charging pulse (10^{-4} sec) the current I_{L3} is not to exceed 20% of the total available current I_3 (0.2 mA), so that not more than about 10% of the current is wasted; dI_{L3}/dt must then be not less than:

$$\frac{0.04 \times 10^{-3}}{10^{-4}} = 0.4 \text{ A/sec,}$$

hence L_3 should be greater than $315/0.4 \approx 800 \text{ H}$.

3) J. F. H. M. Stumpers, Philips Res. Rep. 5, 270-281, 1950.

The meter circuit

As explained on p. 369 the mean anode current flowing in the pentode *P* is a measure of the average number of counts performed by the counter tube per unit of time. The most obvious method of measuring the pulse frequency is therefore to connect a meter with a series resistor, in parallel with a capacitor in the anode circuit of *P*. Both the meter and the series resistor then occasion a certain voltage drop, however, necessitating a battery *B*₁ of higher voltage and involving a higher battery drain by the whole unit.

It is found that a lower consumption of current can be achieved by means of another circuit, as shown in fig. 5. Here the milliammeter *M* with resistor *R*_m

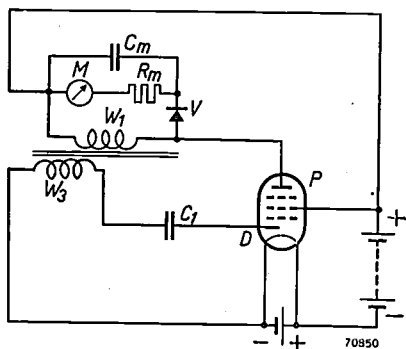


Fig. 5. A method of connecting the meter which is an improvement on a meter placed directly in the anode circuit of the pentode.

is in parallel with a capacitor *C*_m, this circuit, together with a rectifier *V*, being connected in parallel with the primary winding *W*₁ of the transformer. Each time that the pentode *P* passes a current pulse, a damped oscillation occurs in the oscillatory circuit formed by winding *W*₁ and its own self-capacitance. Now, the polarity of the rectifier element *V* and diode *D* is such that, in the first half of the first cycle, the capacitor *C*₁ is charged, and, in the second half of the same cycle, the capacitor *C*_m. During the complete cycle effective use is thus made of almost all the available energy.

Apart from this more effective utilization of the energy, this circuit also ensures greater stability of the whole instrument than if the meter were included directly in the anode circuit of *P*, for, if the oscillation in *W*₁ were allowed to decay freely, there would be some risk that the second period would assume the function of a count. The pentode would thus be triggered, whereupon the process would be repeated, and the number of anode current pulses would become independent of the number of counts. By contrast, if practically all the energy is utilized

during the first period in the manner just described, the amplitude of the decaying oscillation is very small and all risk of the above-mentioned undesirable consequences is eliminated.

The circuit consisting of the diode-pentode may be regarded as a D.C. transformer fulfilling the same function as the vibrator unit in car-radio sets. The voltage supplied by the battery *B*₁ is "transformed" up to the necessary value for the counter tube; this voltage is mainly dependent on the battery voltage and the "transformation ratio" and adjusts itself automatically to roughly the required value, irrespective of the number of pulses per second. Fig. 6 shows the results of measurements that confirm this. The current flowing in the counter tube and the diode are each plotted as a function of the voltage on the counter tube (in this case externally applied), for different numbers of counts. The voltage on the tube of course adjusts itself to the value at which these currents are equal. It is seen from fig. 6 that for a variation in the number of counts from 50 to 600 the voltage on the counter tube varies only between 385 and 360 V.

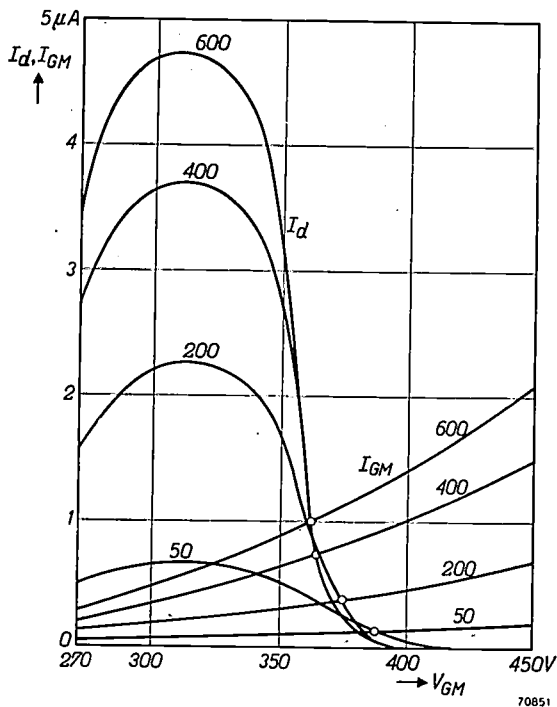


Fig. 6. The diode current *I*_d and counter tube current *I*_{GM} as a function of the (externally applied) voltage *V*_{GM} on the counter tube. The values shown on the curves indicate the number of counts per second. The voltage across the counter tube adjusts itself to the point of intersection of the relevant curves; this voltage is seen to be only slightly dependent on the number of pulses per second.

Final design of the apparatus

The circuit in the ultimate design differs from the diagrams given above in some respects; this final

The feed-back necessary for the pentode to supply anode current pulses, is produced by a capacitor C_3 . It could also be obtained without this capacitor, viz. across C_1 , R_1 and C_2 (if R_1' and C_1' were not provided), but it was found that a capacitor connected in the manner shown ensures shorter duration of the anode pulses, this being advantageous when the number of counts per unit of time is high. In order to avoid feed-back across C_1 , R_1 and C_2 as well, a resistor R_1' and capacitor C_1' are included.

It is very important that the duration and wave-form of the anode current pulses shall be as constant as possible in all circumstances, for these determine the calibration of the meter. The latter is affected considerably by the voltage supplied by the anode battery B_1 ; since this voltage drops during the life of the battery, two resistors R_b and R_b' are included for switching in series with the battery for "high" and "average" battery output. The values of these resistors are such that on "normal" battery output a voltage-drop occurs which is just sufficient to reduce the anode and screen grid supplies to the required values during the pulse. The strength and duration of the pulses are thus rendered independent of the value of the battery voltage. A capacitor C_4 is connected in parallel with B_1 to prevent the increasing internal resistance that accompanies ageing of the battery B_1 from affecting the pulse duration.

The instrument is switched on by momentarily closing the switch S_1 ; a positive voltage, obtained from the potential divider R_3 - R_4 across the battery B_1 , then compensates the voltage from the battery B_2 .

When the number of ionising particles entering the counter tube per unit of time is too small to be indicated by the instrument, the current pulses can be rendered audible by means of headphones; the terminals provided for this purpose are denoted by T_1 in the circuit diagram ⁴⁾.

Fig. 9 depicts the Geiger counter in use and fig. 10 shows the instrument with cover removed. The weight is 520 g (approx. 18 ozs.) and the dimensions are $17 \times 4 \times 10$ cm (approx. $7'' \times 2'' \times 4''$). Small batteries of the kind supplied with hearing-aids are used. A knob with 6 positions is provided on the front panel for operating the switch S_1 , the sensitivity switch S_m and the filament heater switch S_f . The position of the knob at which S_1 is closed is immediately next to the "off" position.

⁴⁾ In exploration work it is often more convenient to use headphones, as the eyes do not then have to be kept on the meter.

In general, rotation of the switch through the "start" position to one of the other settings is sufficient to put the instrument into operation. The meter can also be employed to check the voltages of the filament battery B_3 and the H.T. battery B_1 , these operations being also effected by rotating the knob to the appropriate positions;



67065

Fig. 9. The Geiger counter in use. The illustration reveals the fact that the very weak radio-active radiations from the luminous dial of a wrist watch can be registered.

the associated circuits are omitted in fig. 7 for the sake of clarity. Resistors R_b and R_b' are placed in circuit with the aid of a separate switch, and two marks on the meter indicate the anode voltages at which this switch should be set, in accordance with similar symbols.

Calibration of the instrument

As already mentioned, the instrument has two scales, one for 0-40 and one for 0-600 counts per sec. Calibration of these scales would appear to be simpler than it actually is; it might be supposed that it is only necessary to connect a counting device to the instrument in order to be able to count the number of pulses per second for each deflection of the meter. For a sufficiently small number of counts this would certainly be possible, but, with an increase in the number of ionising particles, the number of discharges in the counter

tube drops below the number of particles entering, seeing that the "dead time" of the counter tube as well as that of the blocking oscillator then become more and more significant (see article referred to in footnote ¹). If this dead time were known, it would be a fairly simple matter to compute the number of ionizing particles from the number of discharges per second, but the dead time is not

meter. The dead time of these instruments is so short that it is of no significance in the measurements. The number of pulses per second can be controlled by varying the distance between the radio-active substance and the scintillation counter, thus providing a ready means of calibrating the Geiger counter.

The fact that only the number of "pulses" of

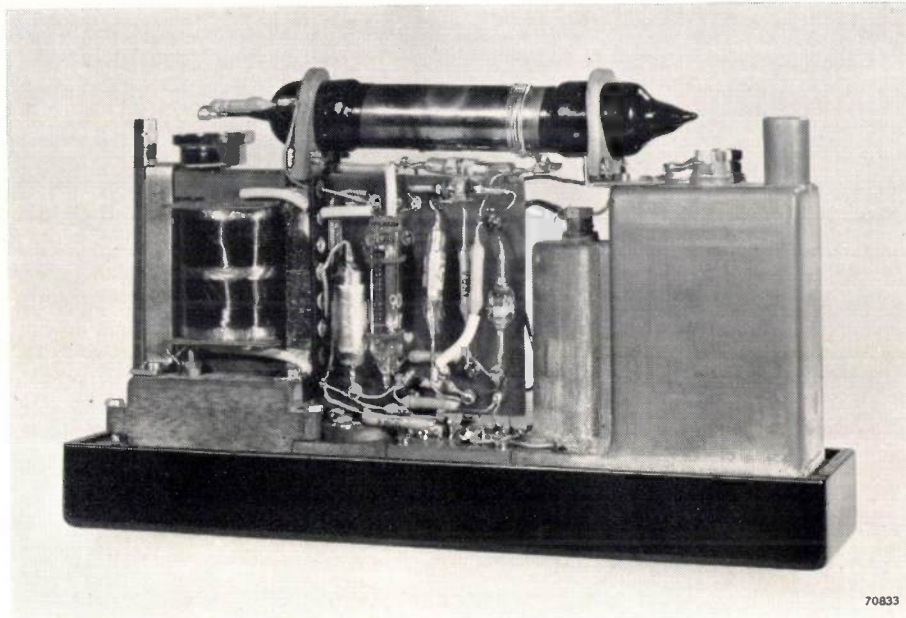


Fig. 10. The instrument with cover removed. At the top will be seen the counter tube. At left the transformer, at right the filament and H. T. batteries.

usually known. More accurate calibration is ensured, therefore, by measuring the number of ionising particles with an instrument whose dead time is negligibly short.

By the method actually adopted use is not made of individual particles or photons, but of pulsed X-rays, the pulses being of such short duration that each produces only one count, although the number of photons per pulse is very high.

The Geiger counter is placed in the vicinity of an X-ray tube fitted with a grid which is biased so that normally no anode current flows: positive pulses applied to this grid result in anode current, thus producing pulsed radiation. The frequency of the voltage pulses for the grid varies arbitrarily, the pulses being derived from a radio-active substance emitting α -particles, placed in the neighbourhood of a scintillation counter. In this apparatus voltage pulses are produced each time an α -particle strikes a zinc sulphide layer mounted in the unit; these pulses are amplified and then applied to the grid of the X-ray tube, as well as to a "scaler" which counts the number of pulses, or to a counting-rate

X-rays (or of artificially produced radiations in general) emitted for very short periods of time is counted, and not their intensity, is the main reason why calibration in roentgens per hour is not provided. In certain cases it would otherwise be possible to conclude quite wrongly from the deflection that a given radiation were harmless. Another reason why it is not practicable to calibrate in roentgens per hour is that the sensitivity of counter tubes to γ -rays is dependent on the particular wavelength of these rays.

Summary. In the design of small portable Geiger counter instruments a low current consumption is of considerable importance; the smaller types of dry battery can then be used. This is achieved very successfully if the voltage for the counter tube is supplied by a blocking oscillator delivering voltage pulses which are employed to charge a capacitor in conjunction with a diode. The blocking oscillator functions only when a count in the Geiger tube withdraws some of the charge from the capacitor. An instrument has been constructed along these lines, measuring only $17 \times 4 \times 10$ cm (approx. $7'' \times 2'' \times 4''$) and weighing not more than 520 grammes (approx. 18 ozs.), the batteries being of the kind employed in hearing-aids. Some of the requirements to be met by the circuit components are discussed, and the article also describes the circuit associated with the measuring instrument, which is calibrated in terms of the number of ionizing particles reaching the counter tube. The circuit is specially designed to ensure low battery drain.